DOCUMENT RESUME

ED 162 108 08 CE 018 223

AUTHOR	Mitchell, A. M.; And Others
TITLE	Ways to Evaluate Different Types of Career Education Activities: A Handbook of Evaluation Models.
INSTITUTION	American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY	Office of Career Education (DHEW/OE), Washington, D.C.; Office of Education (DHEW), Washington, D.C. Office of Planning, Budgeting, and Evaluation.
REPORT NO	AIR-66700-7/78-TR3
PUB DATE	31 Jul 78
CONTRACT	300-77-0303
NOTE	228p.; For related documents see CE 018 212-222 and CE 018 823
EDRS PRICE	MF-$0.83 HC-$12.71 Plus Postage.
DESCRIPTORS	Administrator Guides; Annotated Bibliographies; *Career Education; *Case Studies; Curriculum Development; Demonstration Projects; Educational Facilities; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; *Evaluation Needs; Faculty Development; *Guidelines; Instructional Materials; Intervention; Learning Activities; Models; Program Descriptions; Program Development; Program Effectiveness; *Program Evaluation
IDENTIFIERS	Education Amendments 1974

ABSTRACT

This evaluation handbook was developed as a part of a study conducted to identify and describe evaluated, exemplary activities in career education (K-12) which represent the best of the current career education programs and practices referred to in Public Law 93-380. (See CE 018 212 for the final report of this study.) An introductory section discusses such questions as Why is this document needed?, What can evaluation do for your activity?, How do you decide whether or not to evaluate?, and Can you avoid common problems in evaluation? Then the following six evaluation models are described: (1) evaluation of supplementary activities; (2) evaluation of career education facilities; (3) evaluation of indirect interventions; (4) evaluation of staff development; (5) evaluation of product development; and (6) evaluation of curriculum implementation. A general overview is given for each evaluation model and is followed by such information as rationale and perspective, general types of questions to be answered, and steps in implementing the model. Each model description concludes with a case study for analysis. An annotated bibliography with selected references for all six models is also included. Overviews of the ten exemplary projects selected in the study are appended. (BM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED162108

Ways to Evaluate Different Types of Career Education Activities:

A Handbook of Evaluation Models

A. M. Mitchell

With Models by

A. B. Chalupsky
O. V. Baker, J. A. Hamilton, and D. H. McLaughlin
G. B. Jones
S. M. Jung
C. B. Kaplan and P. R. Dahl
A. M. Mitchell

Illustrations by Susanne Baker

American Institutes for Research
in the Behavioral Sciences

Palo Alto, California

31 July 1978

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The information reported herein was obtained pursuant to Contract No. 300-77-0303 with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to document information according to their observation and professional judgment. Consequently, information, points of view, or opinions stated do not necessarily represent official Office of Education position or policy.

CE 018 223

ACKNOWLEDGEMENTS

The writers of this handbook acknowledge the contributions of their project colleagues, Susan McBain, Dick Bond, Dewey Lipe, Pat Spurr and Beverly Harlan, and of the external advisory panel members, Bob Adkison, John Bowers and Art Duffy.

Dr. Alice Y. Scates, from the Office of Planning, Budgeting and Evaluation, who served as USOE Project Officer, was especially helpful. Her interest, suggestions, and support are gratefully acknowledged.

TABLE OF CONTENTS

Rationale 1

Why Is This Document Needed? 1

What Can Evaluation Do For Your Activity? 3

How Do You Decide Whether or Not to Evaluate? 6

Can You Avoid Common Problems in Evaluation? 8

How Might You Approach Evaluation? 11

Why Six Evaluation Models? 14

Model 1: Evaluation of Supplementary Activities, A. B. Chalupsky 19

Model 2: Evaluation of Career Education Facilities, C. Kaplan & P. Dahl . 40

Model 3: Evaluation of Indirect Interventions, A. Mitchell 63

Model 4: Evaluation of Staff Development, O. Baker, J. Hamilton, D. Mc Laughlin 95

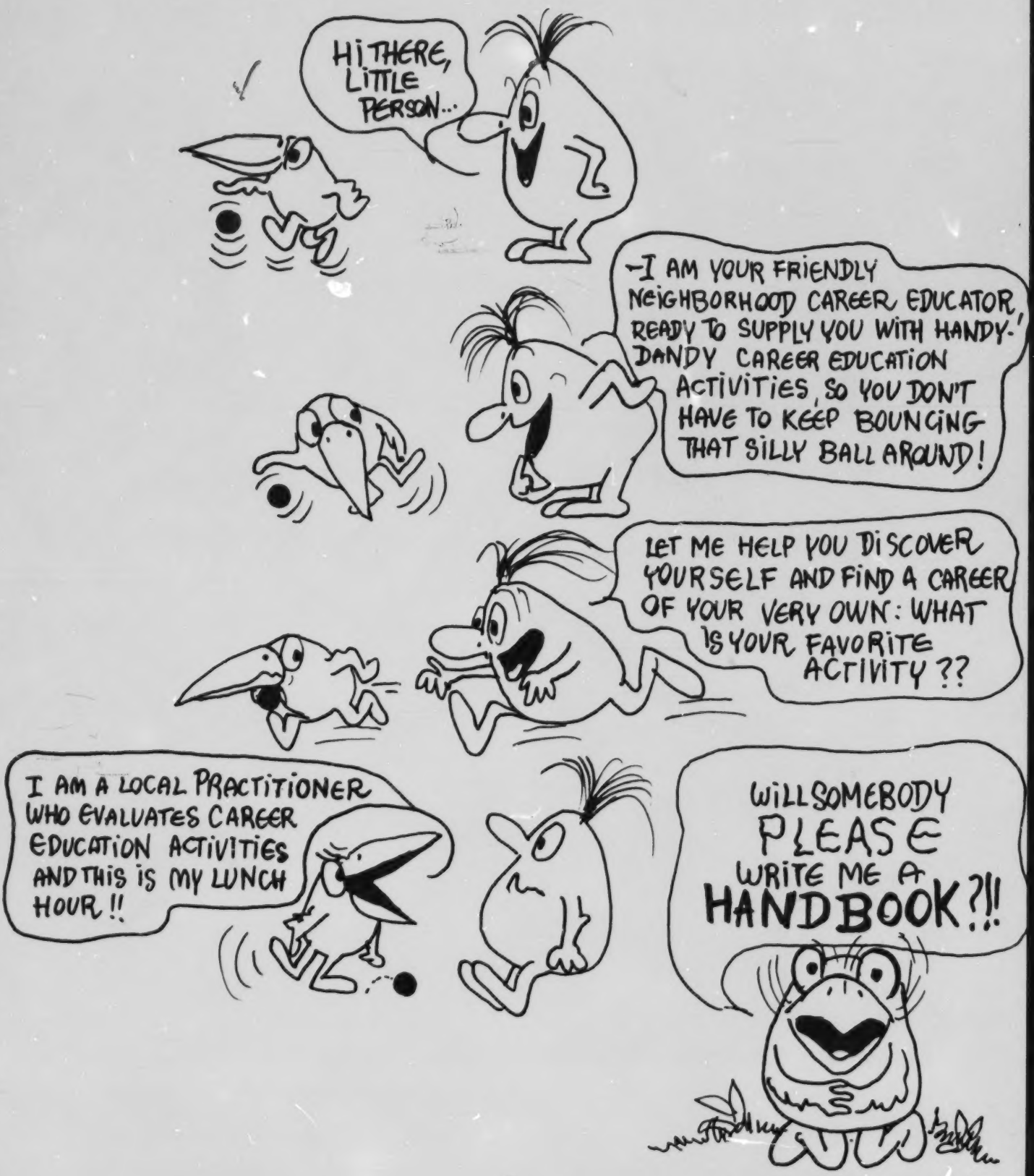
Model 5: Evaluation of Product Development, S. M. Jung 124

Model 6: Evaluation of Curriculum Implementation, G. B. Jones 165

Selected References 202

Appendix A: Ten Evaluated Career Education Activities A-1

Appendix B: Examples of Promising Activity Evaluations B-1


HI THERE, LITTLE PERSON...
-I AM YOUR FRIENDLY NEIGHBORHOOD CAREER EDUCATOR, READY TO SUPPLY YOU WITH HANDY-DANDY CAREER EDUCATION ACTIVITIES, SO YOU DON'T HAVE TO KEEP BOUNCING THAT SILLY BALL AROUND!
LET ME HELP YOU DISCOVER YOURSELF AND FIND A CAREER OF YOUR VERY OWN: WHAT IS YOUR FAVORITE ACTIVITY??
I AM A LOCAL PRACTITIONER WHO EVALUATES CAREER EDUCATION ACTIVITIES AND THIS IS MY LUNCH HOUR!!
WILL SOMEBODY PLEASE WRITE ME A HANDBOOK?!!

Rationale

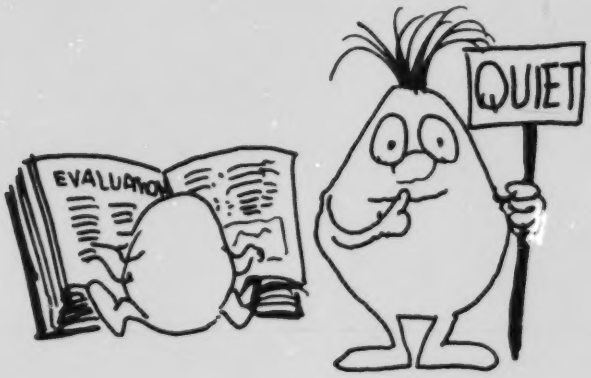
Why is This Document Needed?

What problems does it address? Efforts in career education have been marked by three serious weaknesses: (1) Activities are initiated with little or no examination of underlying premises: the assumptions that are made and

the way these are supported by or conflict with existing theory and empirical evidence. (2) Activities are launched with only the vaguest definition of parameters and no analysis of possible design alternatives, so that each becomes shaped largely by local exigencies and yields little in the way of generalizable information. (3) Questions to be asked and decisions to be made during and after the activity have not been formulated clearly enough to specify data requirements for evaluation. Activity monitoring efforts frequently are spotty or non-existent so that activity managers don't know how to get better results because they don't know what was done poorly.

Who is this handbook designed to help? This Handbook is addressed to local practitioners who are involved in planning, implementing and/or modifying career education activities. Activities include pervasive infusion programs, program components or supplementary activities. The procedures are basically the same regardless of the dimensions of the unit of evaluation.

What are the purposes of the handbook? The purposes of the Handbook are (1) to help you, the practitioner, understand the meaning of evaluation and the importance of designing and implementing a sound evaluation if you do decide to evaluate; (2) to help you decide whether to evaluate your career education activity, and if so, whether to evaluate a single component or all components of the activity; (3) to alert you to common errors or omissions which weaken or destroy evaluation results, and to show you how these can be avoided; and (4) to introduce you to six easy-to-follow evaluation models, each tailored to specific types of career education activities: supplementary activities, facilities development, indirect interventions, staff development, product development, and curriculum implementation.



How can you use this handbook? The Handbook is designed so that each section is self-contained. For instance, even though it would seem repetitious if you were to read all the models, each model is complete in itself so that anyone wishing to implement it will not have to search through other sections to find needed information. However, you are urged to peruse all parts of

this rationale as they will help you determine whether or not to evaluate, what to evaluate, how to avoid common pitfalls in evaluation, and how to choose a model. We have kept this section brief, referring you to other documents for additional information as you feel the need.

What Can Evaluation Do For Your Activity?

What is evaluation? What does it do? Evaluation is a process designed to provide data for management decisions both during and after implementation of an activity. The focus is not solely on proving (product or outcome evaluation) but also on improving, particularly during the first full cycle of implementation. It is not research and it does not necessarily require sophisticated statistical procedures. Its time and resource requirements, its difficulty and its complexity are directly related to the dimensions and complexity of the decisions to be made (e.g., whether to continue to commit resources to the activity) and to the questions that must be answered to make those decisions (e.g. is the activity

significantly enhancing students' career development?). If the decisions are simple and straightforward, based on answers to questions for which relatively simple and striaghtforward information can be gathered, the evaluation will be relatively simple. Examples given in the models illustrate this simplicity.

Rather than an audit or post mortem at the end of an activity, evaluation is a continuous process intended to keep pace with needs and changes. Focus is on the program rather than on the teacher or learner: "Does the program work?" rather than "Did the teacher do a good job?" or "Did the learner succeed?" Evaluation provides a monitoring system which begins operation with the planning stage, provides for feedback during the implementation stage, and encourages mid-course correction to ensure successful achievement of objectives.

A good example of the dynamic functioning of an evaluation system is NASA's monitoring of our space shots. The best human, technological and material resources are assigned to preparation of each space shot; the best geographical spot and atmospheric conditions for collecting desired data are chosen; the target is the one most likely to be reached and to provide the data needed. Before the space vehicle is launched, there is clear specification of information needs, sources of information, how the information will be collected, and what will be done if the information is not transmitted as planned. Alternate strategies are carefully worked out; all possible problems and failures are explored, and contingency plans are outlined. Once the space ship is launched, a sophisticated monitoring system tracks every facet of the planned activity; if the vehicle goes off course, appropriate corrections are made, or alternate plans are initiated. If the vehicle cannot be put back on course or a satisfactory alternate plan initiated, the mission is aborted, rather than spending additional resources on an activity that will not yield the benefits intended.

Applying this concept to your career education activity, you will need to apply your best resources to planning, set reasonable and important objectives, determine the best strategies for reaching the objectives, develop contingency plans, and develop a monitoring system which will be in place before you begin your activity. If your activity is already operating, your evaluation procedures will need to provide for tracing the effectiveness and fidelity of the planning and implementation tasks to identify causes of successes and problems.

Evaluation as a management tool. Evaluation is a necessary component of good management, providing data for decisions. Decisions may relate to extension or continuance of an activity, or to change necessary to make the activity successful. Typical decisions related to each type of activity are presented in the evaluation models which follow this section. Hopefully, as you read these models you will learn about decisions you hadn't thought about before, and learn to ask different questions.

No product or procedure, whether in education or in any other complex field, is completely satisfactory the first time it is created and installed. It must be expected that many modifications will be required before the activity will begin to reach its intended performance level. Each time the activity is repeated, if your evaluation has been planned to answer relevant questions, it should prove to be more effective. In other words, evaluation serves as a management tool, helping you refine your activities so that they become more and more productive. Surely this is a goal worthy of your commitment.

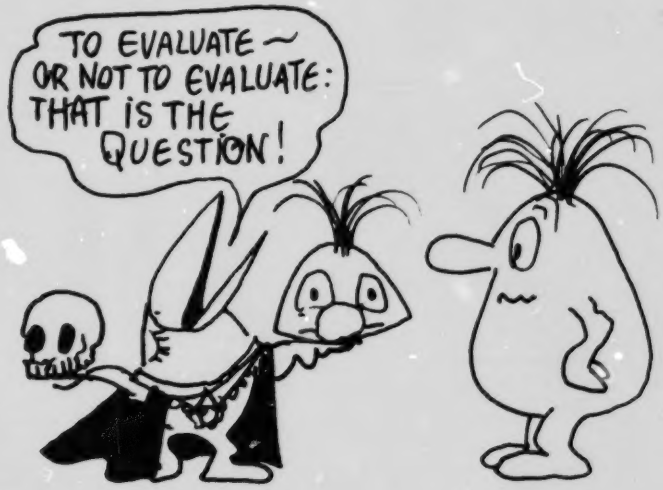
For too long we have repeated activities year after year without taking a real look at whether they made a difference in terms of developing student

competencies. For instance, one school district found that the major outcome of busing large groups of students to a local career fair was that the students ripped off the "most wanted" posters from the FBI booth. There was no evidence of increased career awareness. (This was largely due to the fact that no goals had been established for the activity, and the students had not been prepared to know what to look for.) A good evaluation can help you do better whatever you are doing; doesn't that appeal to you?

Remember that a principal thrust of evaluation is to improve. We can't wait a full year for evaluation results, only to find that an activity is not working. We must sample how it is working all along the way, and plan in-course corrections to make sure we achieve our desired outcomes. Attention to ultimate outcome data should be given only when decisions require such data. The evaluation-revision cycle is operative at all stages of the activity. This is in marked contrast to the myth that evaluation imposes rigid specifications from which you may not deviate.

How Do You Decide Whether or Not to Evaluate?

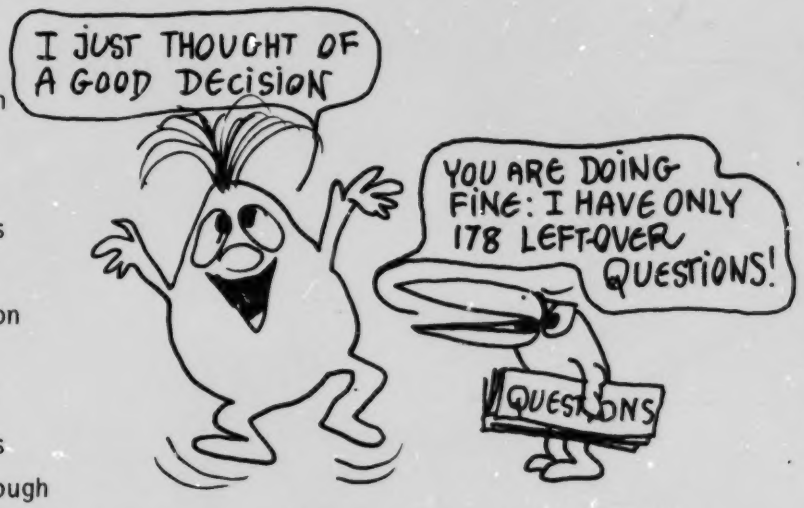
What do you want to find out? It is wasteful of your time and energy and it is unfair to activity planners and implementers and participants, to evaluate an activity if you do not intend to use the information provided by the evaluation. You should not evaluate unless you are committed to



changing the activity, setting priorities, influencing decisions, or justifying funding. An adequate evaluation requires commitment of resources; an inadequate evaluation does not justify the use of any resources. There-

fore you must first decide whether you need to evaluate your career education activity. If you are implementing it for the first time, you probably will want to evaluate it to improve the likelihood that it will have significant impact. If you have evaluated it for two or more years and the results are consistent, you may want to devote your evaluation resources to other activities or at least reduce the scope of your evaluation. If your evaluation has easily observable results with which students, parents and staff are satisfied (e.g., an increase in numbers of students doing volunteer work in the community after a "Know Your Community" activity), you may not need to conduct an evaluation at all. If you are unsure of the effectiveness of your staff development, you may wish to evaluate only that component.

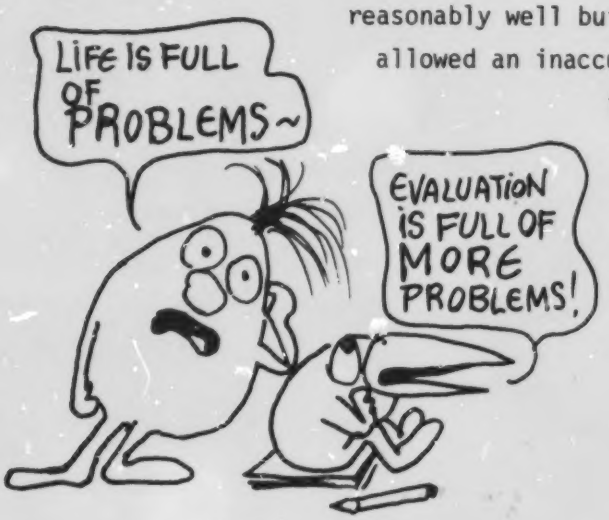
It is easier to ask questions about an activity than to identify decisions that need to be made. Be sure to match your questions with decisions or you'll have surplus questions - or questions in need of a decision. Be sure also that decisions do lead to possible actions. For each decision, with specific alternatives, you need to identify all the rationales (or arguments) that might be used to make the decision one way or another. You will find that most of the rationales make assumptions that need verification through data collection. That is how to arrive at evaluation objectives that will be practical and useful. Properly documented, and with careful follow-through, this step will put an end to the mountains of evaluation reports that have been written and filed, never to affect education in the least.



If you are planning to implement an activity that has shown positive impact when evaluated in another setting, you will have to conduct your own evaluation to see whether it is effective in your setting. You will need to adapt the activity to your context; your staff and your students will differ from those in the demonstration school. These differences need to be addressed in the evaluation.

Can You Avoid Common Problems in Evaluation?

In an analysis of the evaluation of career activities[1], it was found that many project staffs had planned their evaluations reasonably well but, for one reason or another, they allowed an inaccuracy or inconsistency to be introduced that compromised the reliability of the results. Most of these problems could have been avoided with little or no extra cost. A few examples will be given here, along with steps that could have been taken to avoid the problem.[2]



Selection of evaluation sample. A common source of problems is selection of the evaluation sample. Several of the activities reviewed used a treatment/control group design, but failed to establish the comparability of the groups. In some cases the selection procedure introduced obvious bias factors such as using volunteers for the evaluation sample, or having staff members select the groups without reference to any selection criteria. It would have been just as easy to select the groups randomly from total populations, or to establish the relevant variables and locate samples matched on those variables. Another problem with sampling was failure to account for sample losses (e.g., absentees and dropouts). It would have been better to limit the analysis to

[1] Hamilton, J. A., & Mitchell, A. M., Identification of evaluated, exemplary career education activities (K-12). Final Report AIR-66700-7/78-FR. Palo Alto, Ca.: American Institutes for Research, July 1978.

[2] For further examples and more complete coverage of this area, see Bonnet, D. G., The design of career education evaluations in 1975-76. Crawfordsville, Ind.: New Educational Directions, Inc., 1977; and Tallmadge, G. K., The Joint Dissemination Review Panel: Ideabook. Washington, D.C.: NIE/HEW, 1977.

those who participated in both pre and post measurement, provided that attrition did not result in systematic bias between the treatment and control groups.

Selection/development of instruments. Another problem area was selection/development of appropriate instruments for collecting information. Locally developed instruments were seldom supported by satisfactory evidence of reliability and validity, and standardized instruments frequently addressed the activity's objectives only peripherally. Single instruments were sometimes employed to support attainment of two or more objectives, with no subscales available to show differential performance by objective. Success rate on items addressing different objectives should be reported separately to indicate which objectives were achieved and which were not. Whether developing or selecting instruments for your activity, you need to outline a set of specifications which the instruments must meet;[3] if you cannot find instruments which meet those specifications you will need to develop your own. Model 5 in this Handbook suggests guidelines for validating instruments.

Inferring that the treatment caused the gains. Inferring that gains made by the target population were due to the treatment, without ruling out the possibility of other reasons for the gains, was another common failure. This could have been avoided if project staff had made sure there were no intervening variables such as parallel programs in school or community, or unintentional bias in test administration. the evaluator must be alert to possible competing factors and take precautions to see that these factors do not contaminate the data.

---

[3] A good source for locating published measurement instruments that may meet your needs is McCaslin, N. L., and Walker, J. P., Career education measures: A compendium of evaluation instruments. Columbus, Ohio: Ohio State University, Center for Vocational Education, 1978.

Data reduction and analysis. Another problem was failure to group the data appropriately for the analysis that was to be applied. For instance, widely differing subpopulations were grouped together, and the analysis showed no statistically reliable gains; had the data from the subpopulations been analyzed separately, the evaluator might have found that the activity was extremely successful with one or more of these groups. Always be sure that you have chosen an appropriate procedure for analyzing the data, and group your data accordingly.

Establishing comparison standards. Many activities claimed effectiveness without establishing a comparison standard. Pre- and post-testing of the treatment group does not furnish evidence that the activity was the cause of whatever gains are shown. Some type of comparison must be made. Student performance after the activity has been implemented can be compared with that of a group that did not receive the treatment (control group), or with a norm group; trend data and samples of performance over time also furnish comparisons. Standards established by common practice or by expert opinion are acceptable under some circumstances. But some comparison standard must be applied in order to give meaning to the outcome data.

Has anyone been able to avoid these problems? Activities reviewed during the referenced project (see page 6) were limited to those whose evaluation reports were completed by July, 1977. Many had been evaluated in 1975 or 1976. Since that time there has been a tremendous increase in numbers of educators trained in evaluation, in awareness of the need for and the components of a sound evaluation, and in guides, handbooks and models designed to assist the educator in his/her evaluation efforts.[4] The topic bibliography in the appendix of this Handbook will help you find documents that provide detailed descriptions of the issues that concern you, and ways to avoid the problems.

How Might You Approach Evaluation?

How about the creative approach? You probably never thought of evaluation as creative, but as you enter the process and see in it the possibilities for creative decisions which will enhance the quality of the outcomes (and improve staff and student motivation), you will find yourself playing with new ideas. As you pursue the possible consequences of alternative decisions, instead of just letting your activity run itself, you should see your management skills increasing and your efforts valued more highly.

---

[4]Among recent publications are Owens, T. R. and Evans, W. D., Program evaluation skills for busy administrators. Portland, Oregon: Northwest Regional Laboratory, 1977; Capitol Publications, How to evaluate education programs: A monthly guide to methods and ideas that work, Washington, D.C.: Author, 1977-78; and Tallmadge, G. K. and Wood, C. T., User's guide: ESEA Title I evaluation and reporting system. Mountain View, California: RMC Research Corporation, 1976.

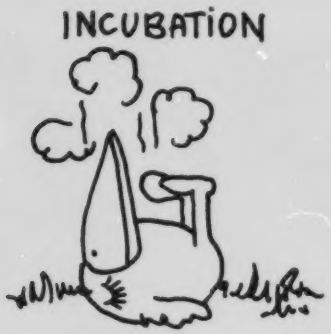
The creative process usually has four stages: preparation (getting the facts); incubation (letting the facts simmer in the subconscious); illumination (receiving spontaneous inspiration); and verification (developing and polishing the creation). Let's apply these to the evaluation process.

• Preparation. This Handbook presents some of the facts about evaluation: reasons for evaluation, what to evaluate, some common pitfalls, and some easy models to follow if you do decide to evaluate your career education activity. Also presented are references to other documents which will expand your knowledge in one or more of the areas touched upon in this Handbook. Use this information to make your why and what decisions.

You need other facts. For instance, you need to think about logistics and staffing. If you decide to evaluate, who will conduct the evaluation? What will be the evaluation schedule? Who will select the evaluation design? Who will select/construct/modify/validate instruments? Who will be responsible for data collection and management? Who will conduct the data analysis and interpretation? After the who questions are answered, you need to address the how and when questions, and the what cost questions (how much money will be required to support the evaluation?). Don't make your decisions about evaluating until you have gotten the facts on these questions. However, don't let the number of questions overwhelm you. If yours is to be a simple evaluation it is very likely that the cost will be minimal (or will be absorbed in the current staff and materials budget) and that there are persons on staff who are able and willing to perform the necessary tasks. Have faith in your own ability to perform at least some of the evaluation tasks.

If, on the other hand, yours is a complex activity and you think you may want to evaluate several components, you may consider the possibility of employing a third-party evaluator on a contractual basis. You might check with your local colleges and universities for names of individuals or groups skilled in the area of evaluation. Perhaps a doctoral student will be interested in helping; such a student may have access to the counsel of the top statistician on the campus, and the cost to you will be minimal. Educational research and development organizations are another source of help. In "getting the facts" don't look only in the usual places for the

usual answers. Be creative and extend the range of your search. A key rule-of-thumb for being creative is to try to see a situation from different angles.



- Incubation. Don't make a quick decision about your evaluation. Let the facts you have gathered simmer in your subconscious, sharpening your intuition. Discuss your ideas with staff members and professional colleagues, and listen. You'll probably pick up some ideas you wouldn't have thought of, and you'll undoubtedly have many of your ideas validated.

- Illumination. One day you'll probably find that you're really turned on to the idea of evaluating your career education activity. You won't know exactly where the turn-on came from, but you'll know you have something worth doing. At this point the facts will fall into place and you will know what your evaluation is going to look like, and you will know that you won't have to look back later and say "If only I had thought of so-and-so." Now you're ready to plan your evaluation.

- Verification. At this point you have decided which component(s) of your activity you will evaluate. Now you'll turn to the models in this Handbook (or in one of the other referenced documents), and choose the one(s) that most nearly fit(s) your needs. Study it carefully, going through each step and filling in the facts from your activity, modifying the model to fit your activity and needs. Where you find some ambiguity or lack of fit in the model, check one or more of

the references, or discuss the issue with a knowledgeable colleague. Be sure to include the career education activity staff in your planning; evaluation, like implementation, will not work well if those who are affected by it are not included in the planning. Remember that all this should take place before the activity is implemented.

During the implementation of the activity, monitor the activity carefully, using evaluation data to make decisions for change; this is when you will feel really creative. Document carefully, include in your decision making all those who will be affected by the decision, and watch the effectiveness of your activity grow. This approach will greatly increase the likelihood that you will have evidence that your activity participants have made significant gains.

Why Six Evaluation Models?

How do you know which model(s) to use? As indicated earlier, models for evaluating educational programs have been developed by a number of evaluation specialists during the last few years. Many of these appeared on the market during this project's operation. At one point we questioned the desirability of adding still more models to the growing literature. But as we studied those that are available we decided that there was room for a set of models organized around specific kinds of activities. We felt that you, the practioner, needed and would appreciate the opportunity to choose a model that directly addresses the component(s) of your activity that you want to evaluate, without wading through decision trees and PERT charts.

As we analyzed the 200+ career education activities that had been submitted as exemplary and evaluated during this project[5], we found that they could be categorized into six activity groups:

Supplementary activities
Career education facilities
Indirect interventions
Staff development
Product development
Curriculum implementation

We studied the ways in which these submitted projects were evaluated and noted that there were differences in approaches appropriate to each activity group. From our analyses of their evaluations we were able to identify techniques as well as problems encountered, and with this experience sharpening our own knowledge of evaluation, we developed the separate models.

You may have some trouble with the term model; so did we. We discussed several other terms that might be used, but none of them was any more descriptive than this one. For our purposes here, a model is simply a step-by-step description of a process for evaluating specific types of career education activities. Each model suggests procedures, instruments, evaluation designs, techniques for data analyses and reporting procedures that have worked for other practitioners evaluating these kinds of activities. There is an underlying pattern in all the models - certain steps that must be pursued in evaluating any type of activity - but sequence, emphasis, and specifics vary. Evaluation is not totally a formula; evaluators, like other professionals, create their own patterns, based on the thrust of the

---

[5]The project referred to is the Identification of Evaluated Exemplary Career Education Activities, funded by the Office of Planning, Budgeting and Evaluation, U.S. Office of Education. See Appendix A.

activity, the decisions and questions to be addressed, and their own perspectives and experiences. Each of these models was developed independently and each reflects the thinking of its author and the differences between and among the types of activities for which they were designed. They may appear to be oversimplified, and perhaps they are. They are based on the assumption that you will have gone through the creative process described earlier before selecting your model(s) and that you will have gathered enough facts (or acquired the services of a specialist) to use these models as guidelines. Their principal value lies in their emphasis on the importance of the interrelationships of the parts of the evaluation process, and on a strong progress-monitoring system.

Which model(s) you select will depend on which components you want to evaluate. If you are planning to install a career center, you may need only to use the Career Facilities Model. If you want only to evaluate the efficacy of your staff development activity, you will need only the Staff Development Model. If you want to evaluate all components of a career infusion program, you may need to use the Curriculum Implementation Model, the Staff Development Model and the Product Development Model. You will find that the models will overlap somewhat, and that the evaluation components of several models can be combined at some points. Look closely at the types of activities each model is designed to evaluate, and you should have no difficulty choosing the appropriate one(s).

You will notice that besides being organized around activities, the models emphasize the decisions that will be made as a result of the evaluation, and the questions that need to be answered to make these decisions. The models show how to address the nine learner goals proposed by the Office of Career Education (Hoyt, 1977), that students should be:

- competent in the basic academic skills required for adaptability in our rapidly changing society
- equipped with good work habits
- equipped with a personally meaningful set of work values that foster in them a desire to work
- equipped with career-decision-making skills, job-hunting skills, and job-getting skills
- equipped with a degree of self-understanding and understanding of educational-vocational opportunities sufficient for making sound career decisions
- aware of means available to them for continuing and recurrent education
- either placed or actively seeking placement in a paid occupation, in further education, or in a vocation consistent with their current career decisions
- actively seeking to find meaning and meaningfulness through work and in productive use of leisure time
- aware of means available to themselves for changing career options - of societal and personal constraints impinging on career alternatives

Development of these models is particularly intended to respond to the provisions of the Career Education Incentive Act (Career Education Incentive Act, 1977) which specified activities which can be funded under this Act and requires sound evaluation of these activities. Fundable activities include:

- instill career education concepts in the classroom
- give inservice training
- carry out career education guidance, counseling, placement and followup
- develop collaboration with handicapped, minority and women's groups and use people from these groups in the classroom and on field trips
- set up work experiences for students who want to explore specific careers, provided there is a chance of jobs in the field and the students don't displace other workers
- hire or train coordinators to pull together career education programs in the LEA or several districts (though not in just one school building)

- conduct institutes for board members and parents, among others
- buy materials and supplies
- operate community career education councils or resource centers
- conduct needs assessments and program evaluations

Since the learner goals and the Career Education Incentive Act encourage such a wide range of career education activities, there is no question of the need for different evaluation models for different activities. The models make it possible for you to evaluate all components of your career education activity, or to limit your evaluation efforts to one or more of the components. This adds a new dimension to evaluation.

If your activity includes a staff development component, a curriculum implementation component, and a product development component (e.g., development of learning activity packages), it would be inappropriate to evaluate student outcomes only. When you evaluate all components, not only are you attempting to evaluate the success of the total activity, you also are sorting out what components contributed the most toward success. Some component evaluations (e.g., product evaluation) would not have to be repeated when the program is replicated. This saves time and money for adapters/adopters. The idea that some program components can be evaluated apart from the program as a whole is a useful idea as it stresses the concept of gathering only that information that is needed for decisions.

Appendix C lists documents containing other models which you may prefer to use. However, you will find that these are general models, and that they do not address specific career education activities. Here for the first time are discrete models for evaluating supplementary activities, education facilities, indirect interventions, staff development, educational products, and curriculum implementation.

You are encouraged to build one of the evaluation models into future project proposals, and to communicate with the Office of Career Education to report your experiences in using one or more of these models; we need to know how they work.

EVALUATION OF SUPPLEMENTARY ACTIVITIES

Model #1

Albert B. Chalupsky

General Overview

Supplementary activities are among the most common examples of the pervasiveness of career education. These activities have the potential for being woven into any part of the regular school program, and, if properly designed and implemented, can contribute to all or most of the goals established for career education. At one and the same time, they can enrich the ongoing career education activities as well as enliven the basic courses in your curriculum.

WE AIM TO ENRICH AND ENLIVEN! GET YOUR SUPPLEMENTARY ACTIVITIES HERE! A TRIP TO HAWAii ~ A NEW CAR ~ ..... WOULD YOU BELIEVE: A TOUR OF OUR PARKING LOT?!
SUPPL. ACT.
SUPPL. ACT
SUPPL. ACTIVITY
SUPPL. ACTIVITY
SUPPL. ACTIVITY
SUPPL. ACTIVITY

Examples of Supplementary Activities

Before going any further, we should be certain that we agree on the answer to the question "What is a supplementary activity?" A concise definition, with which everyone would agree, is probably impossible and not really necessary for this model. Nevertheless, the basic ingredient of a

supplementary activity is that it adds something to the regular program that truly reinforces the day-to-day career education efforts. As compared to other types of career education, supplementary activities are more often of short duration with a definite beginning and ending.

Here are some specific supplementary activities for which this model would be appropriate. You can no doubt think of many others.

- Field trips (e.g., to a food processing plant)
- Hands-on activities (e.g., try-out activities in a ceramics plant)
- Shadowing (e.g. following/observing a bank teller)
- Supplemental curriculum activities (e.g. research/report on a career featured in literature)
- Supplemental guidance activities (e.g., assembly: careers in the '70s)
- Career days
- Career fairs

Why Bother to Evaluate Supplementary Activities?

Whether or not the supplementary activities in your school district supplement and enrich your other career education activities, and whether they justify the time and money expended on them, will never be known unless you identify the specific benefits you expect to achieve and then measure the extent to which these expected benefits are in fact achieved. Don't make the mistake of assuming that just because an activity is considered "supplementary" that it will automatically supplement your regular school program or that just because an activity is fun makes it worthwhile to do. A supplementary activity with a low official budget, or perhaps no separate budget at all, still costs money. Even if it's low-cost, the student and teacher time spent on it might be more profitably directed toward other activities.

A systematic evaluation plan will help you determine, to your satisfaction (and to the satisfaction of those who control your budget), which of these additional activities are truly supplementary and worthy of continuation, or perhaps expansion, and which should be dropped. Of equal importance, a sensible evaluation plan can help you identify what changes should be made to make these activities more effective and, particularly in these days of widespread budget cutting, more cost-effective. Evaluation furnishes data for these and other decisions.

Because of their short-term nature, supplementary activities often escape systematic evaluation. Why evaluate something that is so short and often so inexpensive (at least apparently inexpensive)? Looking at one single career education activity and the cost to design an evaluation program for this one activity, the temptation <u>not</u> to evaluate would be very convincing. On the other hand, if you total up all the resources expended on supplementary activities in your district during the year, it's equally convincing that what is needed is a master evaluation plan or model geared to the features shared by most, if not all, supplementary activities. Parts of this plan can then be routinely applied as a normal part of each supplementary activity, while other parts might be modified slightly with relatively little effort.

All supplementary activities need not be evaluated with equal thoroughness, of course. The extent to which we evaluate will obviously be related to the amount of investment in the activity and how important it is to make it better or to do something else instead. At the very least, however, if you have to defend the doing of a supplementary activity to someone else in order to obtain funds or other support, then you would be wise to determine to what extent it has achieved its purpose and how it could be improved. The results will go a long way toward justifying your next request for support.

There is a most interesting relationship between supplementary activities and evaluation which may have escaped your attention. Without doubt, if you or your staff design a supplementary activity so that it can be evaluated (even if only by answering a few evaluative questions), you can be certain that the supplementary activity will be better than it would have been otherwise, even if you don't actually carry through the evaluation. Naturally, we don't recommend this mythical evaluation as a routine practice. The payoff would certainly be much greater if you not only ask the right questions but obtain the answers and use them to improve your career education activity.

General Types of Questions to be Answered

Outlined below are some of the types of questions you would normally expect to answer at some time or another during your evaluation of supplementary activities. Certainly you will be able to think of additional questions, while on the other hand don't assume that you will have to answer each and every question for each supplementary activity. The questions below are divided into two broad categories: process questions, which cover how you or someone else in your school went about planning and implementing your supplementary activity; and outcome questions, which delve into the type of impact this activity actually had upon your primary target audience, in most cases the students. The outcome questions should be related to the stated objectives established for the supplementary activity.

Process Questions

Each of the following questions can be presumed to begin with the words "To what extent...."

1. Were the stated objectives of the supplementary activity clearly specified and realistic?
2. Were the objectives and the implementation of the supplementary activity integrated with the regular classroom career education activities:
   - to prepare students for the supplementary activity?
   - to follow up the supplementary activity experience?
   - to enhance student motivation for learning basic skills?

3. Were the expected characteristics of the supplementary activity documented in advance of their implementation?
4. Were the resources that were planned for the supplementary activity realistic, in terms of the available budget and time, and adequate to do the job?
5. Were the persons responsible for the supplementary activity properly assigned and adequately trained?
6. Were problems of implementation anticipated and resolved during the design of the supplementary activity?
7. Did unanticipated problems arise during the course of implementing the supplementary activity? How could they have been avoided?
8. Was the planning of the supplementary activity a cooperative effort involving teachers, counselors, parents, and other community resource persons?
9. Was the supplementary activity implemented as planned or, if not, what was the nature of the adaptation?
10. Were the resources (time, money, and materials, etc.) used to implement the supplementary activity efficiently expended and justified in terms of the potential payoff?
11. Were counselors and teachers active participants in the supplementary activities?
12. Did parents and other community resource persons participate in, contribute to, and/or benefit from the supplementary activity?
13. Did business, industry, and labor, professional and community organizations participate in the supplementary activity?
14. Did unanticipated side effects result from the supplementary activity which increased, or detracted from, its value?

Outcome Questions

Here again, you will select only those questions appropriate for your supplementary activity. Each question is presumed to start with the words "To what extent...."

1. Did students gain in occupational knowledge and improve their skills of decision making, job hunting, and/or job getting, as intended by the supplementary activity?
2. Did students gain in knowledge of educational requirements underlying the occupations covered by the supplementary activity?
3. Did students show increased use of career education facilities as a result of the supplementary activity?
4. Did students display increased interest in one or more basic academic courses or in the value of school generally?
5. Did students show more positive attitudes toward work and/or increased desire to work?

6. Were the outcomes worth the cost of the resources expended on the supplementary activity? How does this activity compare in this respect with other supplementary activities?
7. Could the supplementary activity be modified to improve its effectiveness or reduce its cost?
8. Were some students benefited significantly more or less than others who participated in the supplementary activity? What are the resulting implications for modifying the supplementary activity or for improving the assignment of students to this activity?

Steps in Implementing This Model

Described below are the steps to follow in developing and applying an evaluation procedure for a class of similar supplementary activities or, depending on how comprehensive your evaluation plan, for all such activities you expect to implement. Don't be shocked by the number of steps or effort that seems to be involved in their development. Remember, once the procedures are developed you should be able to apply them with little or no modification for each activity you plan to evaluate, using your own good judgment for how thoroughly you apply them. Your judgment should take into account the amount of resources you will be investing in each supplementary activity and the benefits you can anticipate by improving its effectiveness. The specific steps to be covered in this model are as follows:

1. Determine what specific questions, both process and outcome, need to be answered.
2. Specify how the evaluation will be carried out.
3. Estimate the resources, human, technological, and material, required to conduct this evaluation.
4. Develop/adapt/select the appropriate data collection instruments and procedures.
5. Collect the data.
6. Organize, analyze, and interpret the data.
7. Document your findings, conclusions, and recommendations.

These steps should apply generally to any supplementary activity you plan to evaluate. However, we have chosen one type of activity, the field trip, as the common thread to illustrate some of the steps. Field trips represent one of the most popular career education activities outside the school, but one that is often neglected when it comes to evaluation. The

specific field trip we will use for our example will be one for eighth graders to a medical research laboratory. We presume the laboratory is

experienced in conducting such a visit, has provided you with some advance materials, gives an introductory lecture to the students as well as a tour of facilities and one or more demonstrations, and then presents the history of one of the projects by the project director and several staff members. This presentation covers a wide variety of jobs involved from semi-skilled up to professional, and demonstrates the cooperation and teamwork that is necessary to accomplish a research and development project. Following this experience the student group is subdivided into smaller discussion groups.

Step 1: Determine What Specific Questions Need to be Answered

Once you have decided to look at each supplementary activity as a possible candidate for evaluation, the first thing to do is to decide what questions need to be answered in order to provide a realistic and a reasonable evaluation. Earlier in this model, we presented a wide range of questions that can be asked during the evaluation of supplementary activities. Look these questions over again to see which specific ones can be applied directly, or with some modification, to the activities you are planning to evaluate. Remember that not every question will be appropriate for a particular supplementary activity.

You can judge the appropriateness of these questions against the stated objectives of your activity. But don't be surprised if most of the objectives are implied rather than stated, and those that are stated are in need of elaboration in order to do a reasonable evaluation. Accordingly, before you can settle on the specific questions to be answered by your evaluation you will probably have to spend some time making certain that the specific objectives for each supplementary activity are stated clearly and simply. Let's look at a few of the questions raised earlier and apply them to our field trip example. Of first priority is the outcome you expect to achieve from this field trip, as represented by the following questions:

1. Did the students gain in their knowledge of medical research occupations, their working conditions, and their educational requirements?
2. Did students show increased use of those career education resources which relate to the types of occupations covered in the field trip?
3. Did students display increased interest in math and science courses following the field trip?

Next, we can look at how the field trip was conducted and how the process could be made more effective, with questions such as the following:

1. Were the objectives of the field trip clearly specified in advance, realistic, and shared by the students?
2. Were the resources planned for the field trip adequate to accomplish the job?
3. Was the field trip integrated closely with math and science as well as other related courses? Were students prepared in advance to know what to look for and what kinds of information they could obtain? Were class discussions and/or other experiences scheduled following the trip, including the application of the field trip information to the basic math and science courses.
4. Did teachers and counselors gain in their understanding of students and their career education needs as a result of the trip?

Once you have selected the questions that you need to answer and have related them to the stated objectives of your supplementary activity, it is time to verify that these objectives are indeed measurable. Such objectives should include the specific performance or outcome expected, a standard of acceptability, and a description of any important conditions surrounding the objective. For example, the measurable objective for the second outcome question above might read "Following participation in the medical research laboratory field trip, students will relate competencies developed in school subjects to specific scientific and technical occupations." Be sure to

specify the various groups that will be benefiting from the activity and the expected benefits for each group. Don't forget that even though students are the major target group, supplementary activities in career education can and should benefit teachers and counselors and, at times, parents and community representatives. Teachers and counselors participating in the field trip may learn about occupations to which competencies developed in their subjects relate. Community/business representatives may learn about student interest patterns, and limitations of educational programs.

Step 2: Specify How the Evaluation Will Be Carried Out

As soon as you identify the questions you would like to answer with your evaluation, you should be able to sketch out a realistic plan for answering these questions. Basically, this means determining how the information will be obtained and how it will be analyzed. By now, you may be saying "Why all this work to evaluate a supplementary activity that may take only a half-day or less?" As we mentioned earlier, plans to evaluate a supplementary activity, such as those covered in this step, should be applicable to many similar activities for years to come. Once developed, it will be necessary only to apply one or more data collection and analysis procedures (modified slightly, if necessary) to each supplementary activity you are evaluating.

Now back to our field trip example. Let's look at the general approaches you can use to answer the types of questions in our example of the field trip to a research laboratory.

Outcome Questions	Possible Approaches
1. Student gains in occupational knowledge?	1. Short test based on material from field trip site
2. Increased use of career education facilities/materials?	2a. Equipment/material usage logs b. Systematic observation in the career center
3. Increased interest in trip-related courses?	3a. Nature and extent of discussion on trip-related math/science topics b. Amount of student absence/tardiness during trip-related course units

Process Questions	Possible Approaches
1. Clear, realistic objectives:	1. Inspection of trip plans
2. Adequate resources planned?	2. Review of planned versus actual expenditures
3. Integration with regular classes?	3a. Recording of selected pre- and post-trip class discussions
	b. Review of selected pre- and post-trip lesson plans
4. Teacher/counselor understanding of student needs?	4a. Post-trip discussion among teachers and counselors
	b. Teacher/counselor survey

As you probably know, there are many different designs you could possibly use to evaluate supplementary activities. Considering the limited nature of these activities, however, elaborate designs are rarely possible and often not realistic. For example, ideally, a control or comparison group would be used in an evaluation where we want to be reasonably certain that a particular outcome is the result of some educational activity or "treatment" rather than the result of some other influence. However, for most, if not all, supplementary activities, we are assuming that the efforts involved in forming a legitimate control group are just not warranted or not feasible, given the restrictions in time and money. We may never be certain that the observed effects are due to the supplementary activity. However, because these activities are typically of such short duration and because the effects you will be looking for are rather unique, the likelihood that the outcomes you measure are the result of other factors seems rather remote. This is very different from the problem you would face if you were trying to measure reading gains that might be attributed to a special project. Here, you would be confronted by many alternative explanations for such gains. Fortunately, in evaluating the impact of supplementary activities, there will very likely be few alternative explanations for most of the changes you will be looking for. Nevertheless. whenever a control group is lacking, it will pay you to verify that the students involved in your supplementary activity haven't been exposed to some other source of similar information, e.g., another course or a TV program on the same topic that could influence whatever it is you are trying to measure.

...AND TOMORROW AFTER WATCHING THE TV PROGRAM FILMED AT A MEDICAL RESEARCH LABORATORY, WE SHALL HEAR A LECTURE ON GOVERNMENT RESEARCH LABORATORIES..
?!!

On the other hand, in order to answer the question "How much did students change as a result of this supplementary activity?" you will need to obtain some indication of "pre-treatment" status. This may or may not mean administering some data collection form prior to the supplementary activity. Conceivably, some data may already exist that could be used to infer the state of affairs before the activity, for example, the material usage logs in the career center. In other cases, such as the measurement of student or teacher attitudes toward the supplementary activity, only "post-treatment" data may be needed.

Step 3: Estimate the Resources Required to Conduct This Evaluation

Once you have specified the questions you want to answer about each supplementary activity and have sketched out the general evaluation methods, it is about time to check just to make certain that your evaluation plans do indeed match your resources. In justifying the resources to be used be sure to distinguish between those needed to develop the evaluation program and those needed to implement it. Remember, the key to a realistic evaluation is to be able to offset the cost of your evaluation against the gains you expect as a result of the improvement in your career education, or the savings you anticipate by making your career education more efficient. Again keep in mind that the resources used to develop your plans and procedures can be

amortized against all of the supplementary activities affected by these plans during at least the next few years. On the other hand, the expenses in administering one or more instruments to evaluate a single supplementary activity must be judged in terms of the cost and potential benefit of this one specific activity. It is obvious, then, that the key to a sensible evaluation system for supplementary activities is to design one with a great deal of flexibility and a number of different parts that can be administered, scored, and analyzed very inexpensively. How this will affect the design of specific instruments is touched on next.

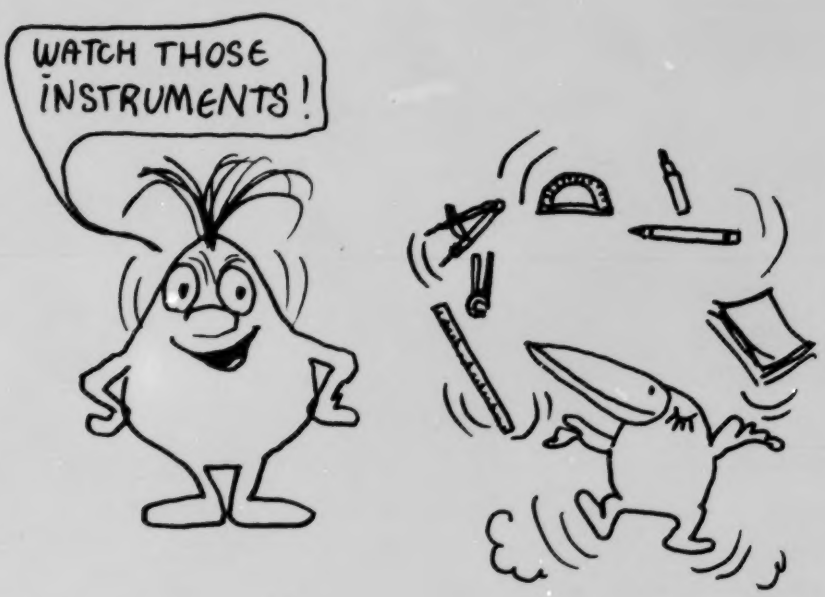
Step 4: Develop/Adapt/Select Appropriate Data Collection Instruments and Procedures

Because of the rather unique objective of your supplementary activity and the fact that only modest benefits can be expected as a result, chances are you will have to develop or adapt most, if not all, of the measuring instruments to be used. Even then, it may often be difficult to detect improvements unless you keep this problem in mind when you are developing your procedures. Don't expect a major or general impact because of one field trip, for example. What you will need to measure are those specific effects that you can reasonably expect to occur as a result of this one experience.

Also keep in mind the importance of measuring whatever you decide to measure shortly before the trip and as soon as possible afterwards, so that you can be justified in concluding that whatever effects did occur were very likely due to this supplementary activity rather than to other causes such as the overall career education program.

When you think of measuring program effects the first thing that comes to mind are tests, questionnaires, and interviews. While these techniques are certainly the most frequently used, there is a whole class of other techniques called unobtrusive or non-reactive measures which you may want to consider in developing your evaluation plans for supplementary activities. Oftentimes, these techniques are used because, as the name implies, they don't disturb the activity being assessed. In the case of evaluating supplementary activities, however, some of these techniques may be especially appropriate because they are convenient and economical.

Basically, there are three kinds of non-reactive measures: physical traces, records, and unobtrusive observations. Cigarette butts are one example of physical traces that could be collected in the student smoking area before and after a school campaign against smoking, in order to measure its short-term effects on student behavior. An even handier type of non-reactive measure are records that are already maintained, for example, student absence and tardiness records; or records that could be maintained with little additional effort, for example, a student attendance log at a career education booth. A third type of non-reactive measure, unobtrusive observation, is typified by the person who inconspicuously wanders around during a field trip, noting the number of students asking serious questions about each occupation displayed. Other observation methods can also be used profitably in career education evaluation, but as the observer becomes more obtrusive be sure to take into account the potential effect of this observer on the activity or process being assessed. If you feel you would like to try one or more types of unobtrusive measures in your evaluation program, you would be wise to read the book by Webb, Campbell, Schwartz, and Sechrest (1966).



Step 5: Collect the Data

Depending on the completeness of your supplementary activity evaluation kit and the appropriateness of the specific instruments you choose to apply to a particular activity, the actual collection of the data should be simple, quick, and inexpensive. If it is not, something went wrong that needs fixing.

In most cases, the data collection instruments to be used should be well-structured and self-explanatory, requiring only a minimum of training on the part of the data collector. Nevertheless, even highly structured instruments can be influenced by the manner in which they are presented to the students. To reduce the danger of contaminating the data, you would be wise to have the data collected only by those individuals whose job or career progress are not dependent on the evaluation results. Where the data are more subjective and where judgment is required on the part of the data collector you should make certain that the individuals selected to collect the data are sufficiently trained to use the instruments you have developed, have appropriate guidelines for use on the job, and are periodically monitored during the data collection period.

Typically, the data collection phase will be of short duration in order to tap the immediate effects of the supplementary activity. Therefore, you may need to utilize a group of data collectors for a short period, such as one day or less. Depending on your own situation and the regulations of your school district, it may be possible to use advanced students to collect the data as part of a class or extra-credit project.

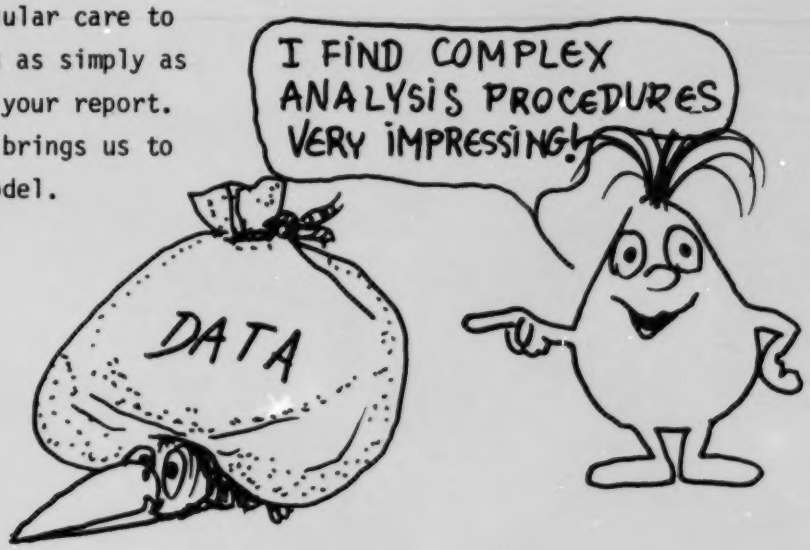
Step 6: Organize, Analyze, and Interpret the Data

A cardinal rule in data analysis planning is to make certain that the sophistication of the analysis procedures matches the information needs, on the one hand, and the quality and amount of data on the other. Still another factor to consider in deciding on particular analytic methods is the level of sophistication of your audience. Judging from the quality and type of data you will most likely be collecting on your supplementary activities, you would be well advised to select the least sophisticated techniques that enable you to obtain answers to your questions. Statistics such as arithmetic means, medians, frequency distributions, percentages, and simple two-way tables, in many cases, will be adequate for your data and appropriate for your audience. Calculating standard deviations will be helpful in your analysis, however, even if you don't necessarily report them in your write-ups.

In appraising the differences between pre- and post-treatment means, t-tests for correlated samples will usually be sufficient. Where your data are nominal, e.g., Yes-No, you can test the differences before and after your supplementary activity using the chi-square technique. Where ordinal data such as rankings are collected, there are a number of non-parametric

or distribution-free statistics, e.g., the signed-ranks test, that can be used. (See Siegel, 1956.)[1] Non-parametric techniques have other advantages in their application to the evaluation of supplementary activities; they are both simple to compute and appropriate for small samples.

If you must use complex analysis procedures, take particular care to explain your procedures as simply as possible to readers of your report. Speaking about reports brings us to the last step in our model.



Step 7: Document Your Findings, Conclusions, and Recommendations

No one will expect you to prepare a major report following the analysis of a particular supplementary activity, such as a field trip. Nevertheless, it will be wise to document your findings and conclusions as soon as possible after each evaluation. Come the end of the year, or the time when you are preparing a request for the funding of your supplementary activities, you will be thankful to have this documentation put together to include in an overall report.

What are the bare essentials that should at least be noted in each supplementary activity evaluation? Our suggestions are outlined below:

---

[1]Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

- Program Description

  Scope of the Supplementary Activity

  - Numbers and kinds of participants
  - Specified objectives

  Personnel

  - Numbers and kinds of staff
  - Staff duties and time involved

  Procedures

  - Organization details
    - period of time covered
    - physical arrangements
  - Program activities
  - Instructional equipment and materials
  - Parent-community involvement
  - Budget

- Evaluation Report

  Objectives Covered in the Evaluation

  Description of Participants

  - How many
  - Ages and grades
  - Characteristics

  Measuring Changes

  - Measures applied
  - How measures were matched to objectives and student capabilities
  - How observers (if any) were specially trained
  - Time lapse between testings

  Presenting Data

  - What data obtained
  - Descriptive statistics (measures of central tendency and dispersion)

Analyzing Data

What analyses conducted

Basis for judging progress of target group

Reporting Findings

Stated in ordinary language

Findings indicate success

Generalizing from findings

- Preparing Recommendations

What Recommendations Can Be Justified on the Basis of the Data?

Other Opinions and Recommendations

Decisions That Can Be Based on Your Evaluation of Supplementary Activities

There are a limited number of decisions that you will probably be making as a result of these evaluations. Here are our suggestions for the major types of decisions:

- Whether each supplementary activity should be
  - continued as is
  - expanded or reduced in scope or frequency, or modified in other ways
  - eliminated or substituted with some other activity

- Whether to search for other supplementary activities for possible adoption or adaptation

- Whether to modify the procedures for recruiting and/or selecting students, teachers, counselors, parents, and community organizations and their representatives for participation in selection of supplementary activities

- How to improve your procedures for planning and implementing supplementary activities

- How to improve your procedures for planning and implementing your <u>evaluation</u> of supplementary activities

THE CASE OF THE ENTERPRISING CAREER EDUCATION COORDINATOR

Executive Summary

Last year, the career education coordinator of Mt. Evermore District barely survived the annual battle of the budget. His funding was increased by 12%, which was not bad considering the cuts made in other parts of the budget. But it was no cause for rejoicing when hopes had been high for at least a 25% increase. The climate seemed right; the potential payoff to the school district from a sizable increase in the career education budget looked very convincing. What went wrong?

Looking back over the penciled corrections to the budget, it was apparent that every item of the budget sailed through untouched but one--supplementary activities. Here is where the superintendent did her slashing, accompanied by one word, "payoff?". Within the supplementary activities budget, one category, field trips, bore the biggest slashing. Why? The answer seemed obvious. Other categories were either blessed with relevant labels or represented unique ideas, obviously innovative. Only field trips seemed like a holdover from the past. Everyone "knew" that field trips were boondoggles, joy-rides, or recreational outings. What could they contribute to career education? Without any backup data, it was impossible to answer this question. Who knows, maybe this time the superintendent was right in squeezing the supplementary activities budget and slashing the number of field trips. On the other hand, maybe it was a mistake. Unless something was done, the career education coordinator, as well as the superintendent, would never really know and every year field trips would more than likely take it on the chin.

But the career education coordinator decided that this year would be different. Backing up the budget requests would be supporting facts and documented results. Field trips may again be cut or they may expand sizably, but whatever way the supplementary activities budget went there would be supporting data to help in making the decision.

There is no reason that all supplementary activites cannot be evaluated systematically and realistically. But the coordinator's priority this year was to devise an evaluation system that would encompass all career education field trips, since that was the most vulnerable part of his program. The process and impact of each trip would be documented and the findings would be used both to

improve the process of career education field trips and to provide overall data to help evaluate this segment of the career education budget. But how can field trips best be evaluated?

Like all good coordinators, the career education coordinator called in his representatives from each school before the school year began and together they established the goals and objectives for all of the career education field trips for the coming year. Many of these goals had never been stated explicitly before. Once the objectives were agreed upon, the next step was to set forth the questions that needed to be answered in order to provide an adequate evaluation of the extent to which each field trip met its objectives, both in terms of student outcomes (product objectives) and in terms of accomplishing specified tasks (process objectives).

Evaluation procedures and various data collection approaches were then defined and refined over the next few meetings, and by the time the field trips began, the schools were ready to initiate their evaluation program.

Before each trip, records were obtained and/or pre-trip measures were administered as a baseline. Shortly after each trip, the same type of data was collected again. While a number of specific questions were developed to be used wherever appropriate, in general, the questions to be answered were: "Has this trip made a difference?" and "How have the students changed as a result of this career education field trip?". From the process standpoint, basic questions included such items as "How realistic were the objectives?", "How adequate were the planned resources?", and "How well were the trips integrated with the regular classes?".

After the first couple of field trips, it was clear that the procedures for planning and implementing career education field trips could be improved. (And they were.) The evaluation program was already beginning to pay off. As the year went on, fewer changes were made to the mechanics of conducting field trips and more attention was devoted to product or impact evaluation.

By year's end, the career education coordinator had accumulated a respectable amount of respectable data, including test results covering student knowledge of occupations, record data showing changes in usage of career education facilities and materials, and objective information on student behaviors and classroom experiences clearly traceable to the field trips, as shown in part on the next page.

Field Trip	% of Class Meeting Standard on Occupational Knowledge Test				% of Class Using Career Ed Facilities and Materials				% of Class Satisfactorily Completing Trip-Related Course			
	Pre	Post	Goal Met	Sig. Level	Prior Week Ave.	Post Trip	Goal Met	Sig. Level	Prior Week Ave.	Post Trip	Goal Met	Sig. Level
1	30	80	yes	**	10	60	yes	**	65	80	yes	--
2	40	50	no	--	30	30	no	--	70	60	no	--
3	30	70	yes	**	30	50	no	*	70	70	no	--
4	35	80	yes	**	40	60	yes	*	75	95	yes	*
5	50	85	yes	**	40	75	yes	**	75	75	no	--

* .05 level
** .01 level

Figure 1. Summary data on field trip evaluation.

In addition to these "hard data," it was clear that the process of field trips had become considerably more systematic and professional in character. Field trip experiences were really supplementing both the ongoing career education activities as well as the regular academic courses. Now it was up to the career education coordinator to present this information concisely and objectively to the superintendent and hope for the best. Regardless of what she did, the career education coordinator felt he had done his part. While there was no doubt that the process of conducting career education field trips had improved over the year, the superintendent could draw her own conclusions from his report and from the impact data he had supplied; and they both could come to a reasonable compromise based on the information supplied.

Some of the career education field trips might be cut for this coming year, some would surely remain, and some types might even be expanded. Chances are he would receive more funds for extending his evaluation program. Who knows, he might even be promoted.

EVALUATION OF CAREER EDUCATION FACILITIES

Model #2

Carol B. Kaplan

Peter R. Dahl

This model is designed to help you evaluate the effectiveness of career education facilities. Such facilities typically include:

(1) A self-contained career center in the school.

(2) A community career center, housed in a community agency.

(3) A mobile van, available to students and/or adults throughout the community on a scheduled basis.

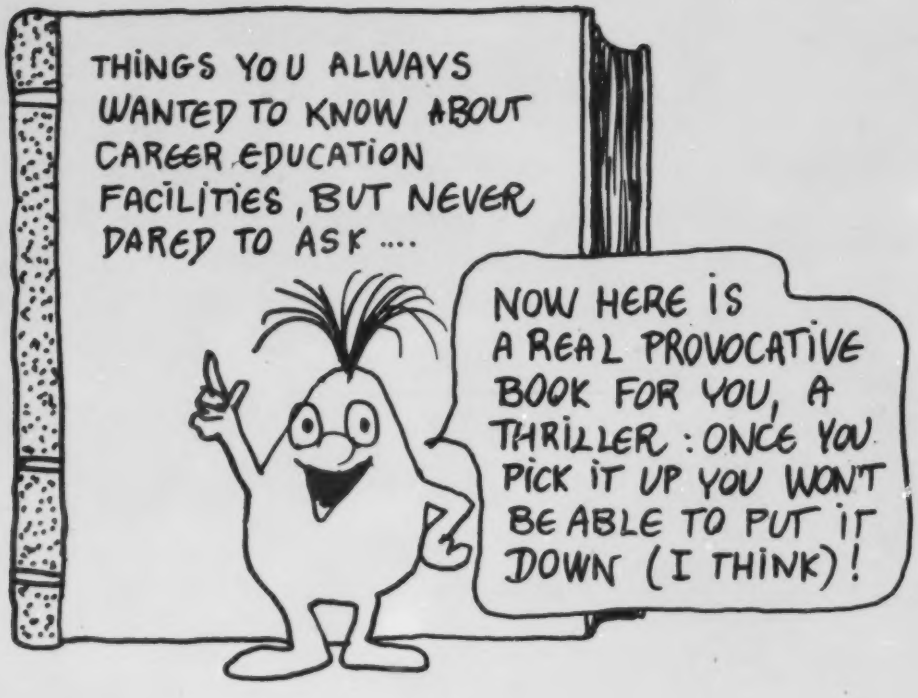
(4) A referral center, such as a telephone dial-a-career-counselor facility.

(5) A career center housed in a section of a classroom.

(6) A section in a library designed to integrate career planning concepts into a school curriculum through a library-based program of career information services.

(7) A career counseling center.

This does not represent a complete list of the configurations a career education facility may present. Your career resource center may be a systemwide coordinating unit for all career education efforts within your district. It may be a center serving all teachers within a state or a county, for the purpose of previewing career education materials and learning how to use them effectively. It may be a center to identify community resources for providing class speakers, field trip sites, job shadowing opportunities for both students and teachers, or curriculum advisors. These are samples of the many activities school-, district-, intermediate-, or state-level educators may develop to meet identified needs.


THINGS YOU ALWAYS WANTED TO KNOW ABOUT CAREER EDUCATION FACILITIES, BUT NEVER DARED TO ASK ....
NOW HERE IS A REAL PROVOCATIVE BOOK FOR YOU, A THRILLER: ONCE YOU PICK IT UP YOU WON'T BE ABLE TO PUT IT DOWN (I THINK)!

Rationale and Perspective

Why Evaluate Career Education Facilities?

Everyone knows that career education facilities are good, so why waste time and money evaluating them? Haven't we seen an unprecedented rush to establish career centers in practically every secondary school? Would school boards support such installations if they weren't good? Yes, they would! Career education is popular with the public, which generally sees it as designed to help students bridge the gap between school and work, and to keep them off the street. So of course a facility that tells them more about occupations has to be good! Innumerable school boards throughout the country have bought career education facilities on faith, but be careful! Your board members may learn about the California Study of Career Centers (Ellis, Ellis, and Mason, 1975). And if they do, they might read it. And if they read it they might ask you for evidence that your facility is contributing to the career development of students. This might involve embarrassing questions like, "How are students changed as a result of the facility?" "What are the objectives of the facility?" "How do you know it's accomplishing these objectives?" "How did you choose the equipment and materials housed in the facility?" "Why? What

are they supposed to accomplish?" "What evidence have you that they are accomplishing the student growth for which they were purchased?" "Are the materials timely, accurate, unbiased, non-distorted?" "What is the relevance of the materials to the interests and needs of the students and the needs of and opportunities in the community?"

Could you answer these questions in such a way that your Board would know that your facility was established after careful consideration of the needs of your students and after careful analysis of the alternate ways those needs could be met? If not, you need this model! The simple truth is that many school districts have established career education facilities because they were popular, because money was available, and/or because they represented an opportunity to show the School Board and the community an immediate accomplishment. There is little evidence that most career education facilities have been carefully planned, designed, and established in response to the needs of students; there is still less statistically reliable evidence of student change as a result of these facilities.

A few relevant findings of the Ellis, Ellis and Mason report include:

(1) Students with no post high school educational plans spent less time in the career center than those who planned to continue their education.

(2) Without ever having gone to a career center, 79% of the students surveyed listed three interest areas; of the students who had gone to a career center seven or more times, 82% listed three interest areas - a difference of only three percent.

(3) Few career centers were planned around student needs, with consistent procedures for assessing needs, determining alternatives, setting priorities, and evaluation.

(4) No significant differences were found between frequent and infrequent career center users on the Planning scale of the Career Maturity Inventory (CMI) (Crites, 1973).

(5) No differences were found between frequent and infrequent career center users on the Information and Decision-Making scales of the CMI.

There are, of course, some positive findings; but the above should alert you to the fact that failure to plan for attainment of priority student outcomes may result in negative or inconclusive results if you do find yourself obliged to evaluate your facility.

The problem with most evaluations of facilities is that they focus on use rather than on outcomes. They identify which students use the facility, which materials students prefer, which teachers assign or bring students to the facility, and -- occasionally -- what tasks or learning packages students complete in the facility. This information is interesting, but it does not provide information that helps make decisions unless it is meticulously related to student outcomes. You need to look at the "if - then" relationships. For instance, "If students use the self-instructional decision-making package as suggested, then these students will be able to outline a sound decision-making procedure." Tying the materials to an outcome helps you identify the necessary focus of your evaluation: what students are able to do as a result of using the facility. The process information (evidence of use of specific equipment or materials), if related to student outcomes, can suggest changes in the physical aspects of the facility, the resources available through the facility, the user services available in the facility, and management procedures such as hours of operation and access to the facility. However, if unrelated to student outcomes, such information doesn't meet the "So what?" test. What difference does it make that students visited the facility X number of times? If those visits did not contribute to measurable/observable growth in their career development, the visits might be considered a waste of precious learning time rather than something to hold up as evidence that your career education facility is "good" and worthy of support.

This model is not intended to discourage you from establishing a career education facility; neither is it intended to make you feel guilty about or to consider closing an existing facility. Rather, it is intended to help you switch your emphasis from materials to outcomes, and your evaluation from laundry lists (lists of materials/equipment available) and head counts (logs of numbers of students using the facility) to evidence of the impact of the facility on the students it is designed to serve. (Of course, what we're saying applies to any educational approach, since the expenditure of resources should be evaluated by gauging impact on students.)

Perhaps you haven't started your facility yet; this model will help you plan it and will help you avoid many costly mistakes in such areas as materials acquisition, equipment purchases, staff deployment, etc. If, on the other hand, you already <u>have</u> such a facility, don't despair! We'll help you sharpen the focus of your efforts and materials so you can prove that it was a neat idea to begin with. O.K.?

Now that we have established the importance of evaluating your facility, let's look at some reasons why facilities may be useful. The 1977 Career Education Incentive Act declares that educational agencies and institutions should make every effort to prepare each individual for a career suitable to that individual's preference. Facilities such as a career information center can help meet this objective by providing information on specific careers and on the preparation they require. A career education facility also can be utilized to help students achieve the following career education learner goals (Hoyt, 1977):

- Equipping students with career decision-making skills, job-hunting skills, and job-getting skills.
- Equipping students with a degree of self-understanding and understanding of educational-vocational opportunities sufficient for making sound career decisions.
- Making students aware of the means available to them for continuing and recurrent education.
- Assuring that students are either placed or actively seeking placement in a paid occupation, in further education, or in a vocation consistent with their current career decisions.

A resource center for teacher use can provide a centralized source of information; it can serve as a convenient central point for the distribution and exchange of instructional materials, as well as a repository of supplementary or background information on career education necessary in the pursuit of prespecified student outcomes.

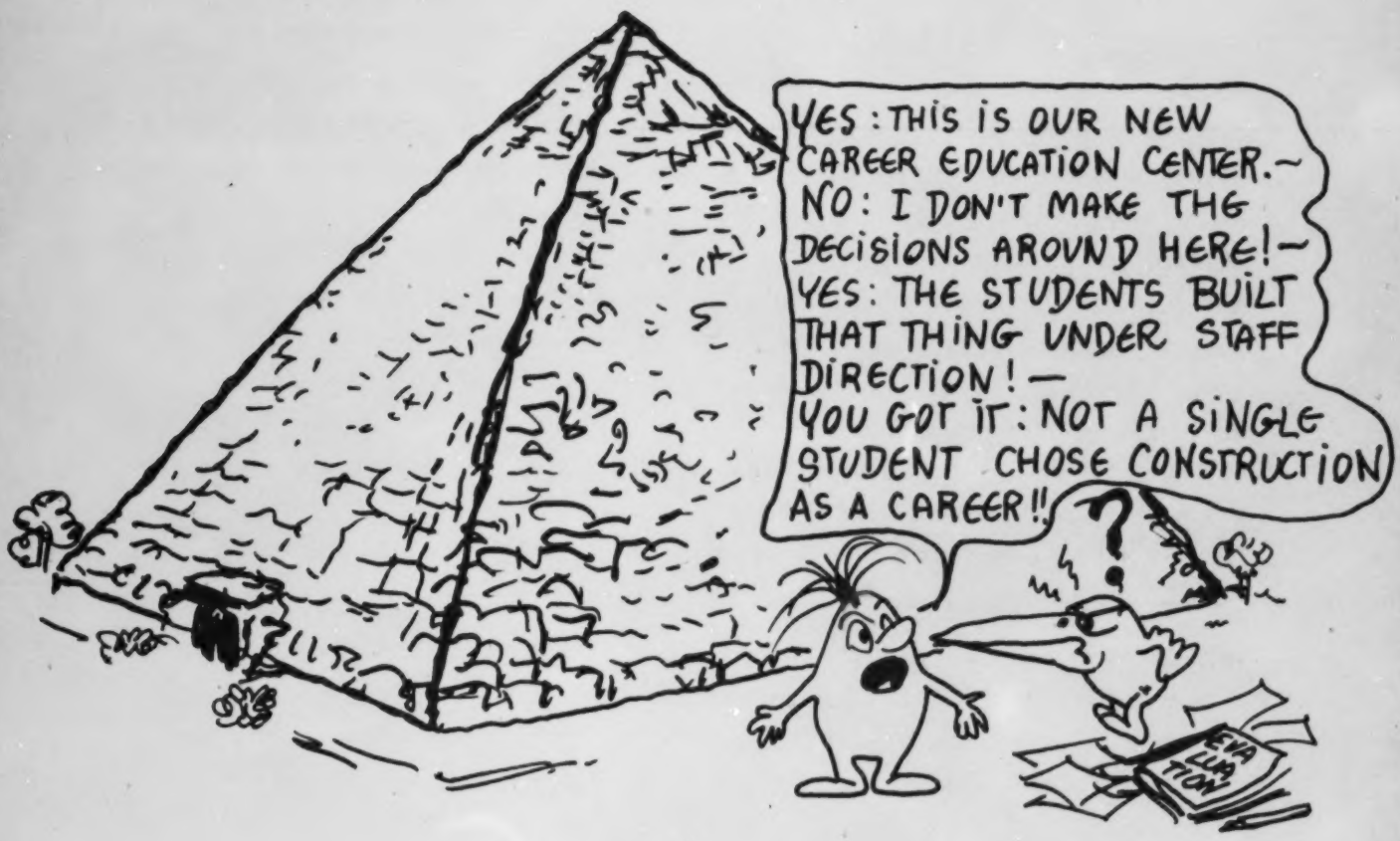
We say a career education facility can do these things; but it will not happen "automagically." You will have to define the goals, design the roadmap, and select the pit-stops along the way if you want your facility to accomplish all those good things you'd like it to accomplish.

Two Kinds of Evaluation

The Rationale and all the Models in this Handbook are intended to impress you, the practitioner, with the importance of planning and conducting a two-pronged evaluation including measurement of processes and measurement of outcomes. Both parts of your evaluation are intended to provide data for decisions: answers to questions that will give you the information needed to make informed decisions.

Process questions include: Were the planned activities implemented with fidelity? Were some of the planned activities unrealistic? Were they unevenly implemented? If so, how did this affect the outcomes? Can the activities be described so they can be replicated? Were timelines established and maintained? Was assignment of responsibilities appropriate? Were some activities or components more effective than others? Were materials effective? Did students use the facilities? For what? How regularly? Was staffing adequate? Were cost projections realistic? These and other process-type management questions need to be answered to provide information for decisions that will ensure the effectiveness and efficiency of your facility's operation.

Outcome questions might include: Do students who complete a planned program in the career center more frequently choose courses of study relevant to their career goals than do students not completing the program? Are students using the career center more able than non-users to identify their strengths and weaknesses, and to relate these to educational and occupational choices? Are students who pursue the career center's planned decision-making program more able than others to demonstrate a consistent decision-making process in real-life situations? Many such questions can be asked, depending on your identification of areas of need for your students.


YES: THIS IS OUR NEW CAREER EDUCATION CENTER.-
NO: I DON'T MAKE THE DECISIONS AROUND HERE!-
YES: THE STUDENTS BUILT THAT THING UNDER STAFF DIRECTION!-
YOU GOT IT: NOT A SINGLE STUDENT CHOSE CONSTRUCTION AS A CAREER!!
EVA LUA TION

From Questions to Decisions

So what's the point in seeking answers to these questions? None, unless you intend to use the information to help make decisions. Be sure to identify all the decision-makers including (a) students, who need to make educational and occupational decisions; (b) teachers and counselors, who make decisions as to appropriate experiences for students; (c) administrators, who have to decide how resources (human, material, and technological) should be deployed; and (d) school board members, who have to decide between and among programs when trying to balance the budget. Institutions of higher education, community groups, and parents may also need information to make decisions that are important to the career development of students. Ask these groups what decisions they need to make, and what information they need so you can design the evaluation to yield that information.

This is an evaluation model, not a planning model, so we will not lead you through the steps necessary to plan a career education facility. However, the decision to establish (or to continue to maintain) a facility differs markedly from decisions to implement other types of career education activities, mainly in terms of costs, difficulties in modifying an established facility, differences in effectiveness of materials and utilization processes, etc. Also, parts of the planning process will be the focus of your evaluation. To highlight the relationship between the planning rationale and evaluation, and to provide perspective on the kinds of questions you may need to address in your evaluation, we need to look at a few planning tasks that have evaluation implications. If you have already established your facility, and you did not perform these tasks prior to implementation, you will need to think them through now; they will help you decide where to place the focus of your evaluation.

You've all heard about assessing needs. Since career education facilities are not only expensive but, once in place, are difficult to change, hide, or abandon, it is imperative that you establish a need for the facility and a focus for its operation before you put an ounce of effort into its design. The need and focus also identify the thrust of your evaluation. You may very well find that the needs you identify can be met more parsimoniously by means other than establishment and operation of a facility. You may even find that a career education facility is contraindicated -- that it is apt to detract from the goal, diffusing students' activities rather than crystallizing them. Unless the needs identified can best be met by the establishment and staffing of a facility, wait until the urge passes, and then go on to something else. It's nice to be able to show visitors a handsome, well-equipped facility, but some taxpayers might get very upset when visiting a pristine facility that appears to be untouched by student hands, or a well-equipped facility being used for rap sessions or socializing or study hall or as a "holding pen" for intractable students.

We will not give space here to telling you how to identify students' career development needs, since so much has been written on the subject already. See the bibliography at the back of this Handbook for appropriate references. But remember that your evaluation will include assessment of the degree to which the facility is meeting the priority needs for which it was established.

After isolating the needs, you should identify indicators that will give evidence that the needs have been met. For instance, if a need is for students to choose courses of study appropriate to their career goals (not for the counselor to make these choices, but for the student to make informed choices based on self knowledge and self direction), an indicator would be a student's projected three-year academic program, with a rationale showing how each course choice will advance him/her toward a specified career goal. Hopefully, you will have selected indicators for each priority need you hope to meet. In some cases a single indicator will do (such as in the above example); in other cases two or more indicators may be required to give a confident picture of the student's need being met. The indicators, student-outcome oriented, are measurable/observable behaviors.

Let's assume that you have sought the most efficient and effective way(s) to help students achieve those behaviors.

- You expanded your thinking: you read about demonstration programs and read journal articles that explained innovative practices; you sought information through your professional organizations, through ERIC, through brainstorming with colleagues. You didn't limit your choices to those strategies rattling around in your own head. You thought big.
- You made a list of all the possibilities, including the possibility of using non-school resources.

I CANNOT MAKE UP MY MIND WITHIN 10 MINUTES, WHETHER I WANT TO BE A HOUSEWIFE AND MOTHER OR PRESIDENT OF THE US. OR A BRAIN SURGEON!!
CAREER EDUCATION FACILITY
I THINK THERE MAY BE SOME LIMITATIONS HERE, LIKE "SPACE", "TIME"...

• You created a reality check by listing real (not imagined) limitations, such as limitations of space, time, personnel, budget, administrative support, etc., that are not open to negotiation. If you were really honest and you performed this task in collaboration with other staff members, you probably were surprised at how many obstacles that you thought were immutable really were subject to change once you established an important rationale for changing them. With your reality checklist, you went back over all the possible strategies for producing the desired student outcomes, and you crossed out those that truly were not feasible.

• Then you ran an effectiveness check on the others. You probably used a rating scale for this, giving a rating of 1 to each strategy that would be sure to produce the desired student behavior, a 2 to each one that probably would produce the behavior, a 3 to the maybes, and a 4 to the unlikelies.

• Next you ran through your list with a rough cost estimate. You tried it on the number "1s" first. You found that one 1 (a career center) would cost about $50,000 for installation and $18,000 a year for staff. Another 1 (a classroom unit) would cost only $20,000 for materials and equipment, and would not require any additional staff. Since the idea was to select the strategy that would be most likely to produce the desired student outcome with the least expenditure of resources, it appeared that the classroom unit, rather than the career center, should be selected. Then you remembered a word of caution: if one were to look for a strategy for meeting a single objective, one probably would never be able to justify establishment of a facility. So you looked at all priority student career development needs, and all the indicators that would signal that the needs had been met, and you found that a well-planned facility, with carefully chosen equipment and materials and with realistic and structured plans for its use, could meet all the needs more efficiently and at less cost than mounting a separate program to meet each need. What has this got to do with evaluation? The indicators you identified will determine the focus of your outcome evaluation.

Let us say that you are convinced (and you've convinced THE POWERS) that the answer to students' career development needs is a career center, housed in a separate room near the library, and it's to be staffed by a paraprofessional who will be responsible for maintenance and use of materials and equipment, for guiding students through packaged career exploration activities, and for cataloguing use of materials and gathering any

data needed for evaluating the facility's efficacy. The regular counselors will schedule themselves into the center one day a week, to be available to students on a drop-in basis.

O.K., you know what it's going to look like--what materials and other resources are required to meet the identified student needs, what indicators will give evidence that the needs have been met--but how are you going to evaluate the facility? Now is the time to contact all those decision makers we mentioned earlier, tell them what you plan to accomplish through the center (in terms of well-defined and well-stated student outcomes), and ask them what decisions they will need to make in relation to the center. The decisions should all relate to the degree to which the center facilitates attainment of the specified student outcomes (outcome evaluation), and the effectiveness of the procedures and resources used to attain these outcomes (process evaluation).

Although process precedes outcomes, we will look first at outcome evaluation; until you have established desired outcomes, you have no benchmarks for determining whether the processes are working. For each step, we will look first at outcome, then at its relationship to process evaluation.

Steps in Evaluating Facilities

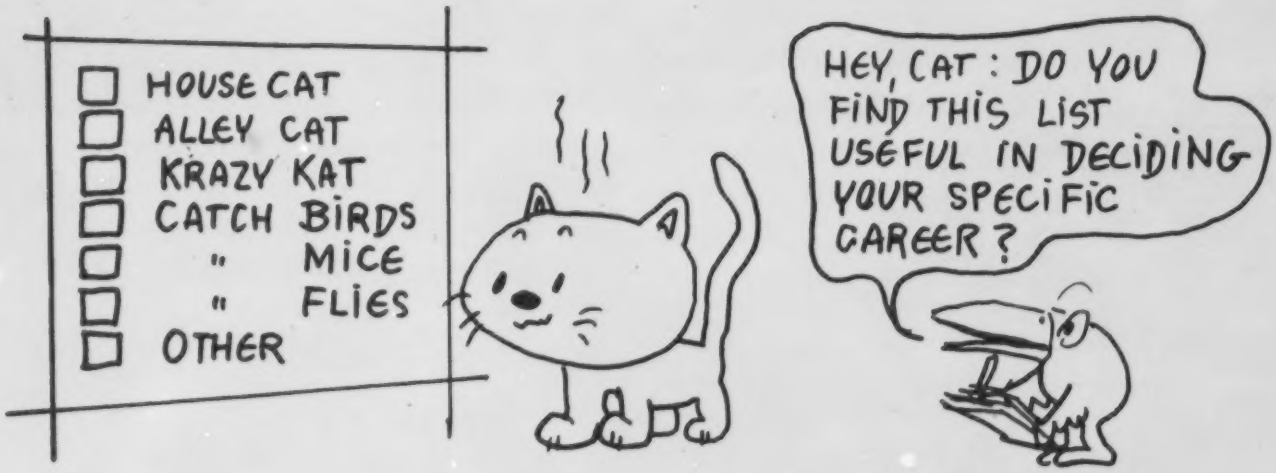
Step 1: Select the Specific Questions You Want to Answer

As stated above, the specific questions to be answered by your evaluation will relate to the problems and needs your facility was designed to meet. There should be a reason for asking each question. When considering whether a particular question is necessary and appropriate, ask yourself whether the answer to the question is needed to make an important decision. An evaluation consisting of "It's nice to know this" type of information will not be worth much in helping you identify ways to make the facility more effective.

Questions will relate to the expected impact of the facility. This is the time you will refer to the indicators discussed on page 48. For instance, if you expected that guided use of the facility would increase students' decision-making skills, you will want to ask such questions as,

"Are students who pursued planned activities in the career center more able than other students to outline and employ sound decision-making strategies?" "Can the improvement in decision-making skills be attributed to use of the career center?" "Are students who used the career center as planned more able than others to project academic programs that will move them toward attainment of stated career goals?" The answer to each of these questions will be an indicator that a need has been met. You should be clear as to how the answers to the questions will affect various decisions. In other words, each question that is asked should pinpoint an information need to test some decision rationale.

For instance, if the answer to the first question is negative, you will need to identify the reasons why the activity did not produce the anticipated results. This suggests process questions, e.g.: "Was the activity implemented as planned?" "Were the materials as effective as anticipated in teaching students decision-making skills?" "Did the staff demonstrate commitment to and expertise in the planned activity?" These questions call for evaluation of the process: a close monitoring and reporting and analysis of the selection, deployment, and use of all resources (human, material, and technological). For instance, the determination of whether or not the activity was implemented as planned would depend on analysis of documentation of task completion. Determination of the effectiveness of materials used with students who participated in the planned decision-making-skill-building activities would depend on documentation of the use of the materials, and of students' demonstrated ability to employ decision making skills before and after using the materials in the planned activities. Determination of the commitment and expertise of the staff could be verified by logs reflecting frequency and extent of implementation of the planned activity, observation of staff performance during implementation, etc. Once you have identified the reasons why the results were not attained as anticipated, you will need to decide whether to modify the activity, secure different materials, abandon the effort, etc. Then focus in on successful attainment of outcomes. What you will do with the answers to your questions--the decisions you will make--constitutes the rationale for asking the questions and for collecting the information to answer them.



Step 2: Decide What You Need to Know to Answer the Questions

The key to completing this step successfully is to determine how to get the information you need while at the same time minimizing the burden on those who will provide the information. To answer outcome questions such as those suggested in Step 1, you will need to specify tasks that students will be asked to perform to demonstrate attainment of the outcome; success criteria for judging the adequacy of students' completion of the tasks will also be specified. For example, using the first question in Step 1, "Are students who pursued planned activities in the career center more able than other students to outline and employ sound decision-making strategies?" you would design a demonstration task such as the following: Given a hypothetical situation requiring an educational decision that will affect future occupational decisions, the student will: (a) outline information needs; (b) relate these needs to relevant values; (c) list alternative choices; (d) list possible consequences of each choice; (e) decide which choice to select; and (f) give a rationale for this selection. Success criteria would include adequacy of information needs listed, relevance of these needs to stated values, range of alternate choices, and relevance of rationale to the selected choice. Depending upon the number of students involved, you might want to have all participating students perform this demonstration task, or you might want to use a random sample of these students. A random sample of students not involved in the activity would also be assessed.

To answer process questions, you would need to plan for documentation of implementation of the career decision-making activity, including: (a) identification of students who successfully completed the activity; (b) criteria used for determination of successful completion of the activity; (c) identification and documentation of unusual circumstances that might have affected student performance; (d) incidence and dimensions of use of materials and equipment; and (e) whatever other factors could illuminate reasons for success or nonsuccess of the activity in terms of student outcomes.

A great help in solving the problem of how to collect the information you need while minimizing the respondent burden is the use of unobtrusive measures. Take advantage of information that already exists to answer some of your questions. Suppose you need to know which materials are seen by students as sufficiently helpful that they access them more than once. You could ask each student to compile a list of the materials s/he found most useful, but this probably would not be as accurate as checking the sign-out sheets to determine how many students accessed each item how many times. The information is available on the lists maintained by the facility. Using available information increases accuracy and saves respondents' time. Avoid putting additional information-gathering procedures into the system if they aren't necessary.

Another example: Suppose you want to determine how optimal use of the career facility can be encouraged. In order to cover all ramifications of use you might develop a questionnaire a centimeter thick. This strategy for collecting information, though comprehensive, is not practical. Your respondents will not complete the questionnaire, and your evaluation will be ineffective. A better approach - one that does not put the burden on those who use the facility, is to keep a sign-in or sign-out sheet with spaces for the user to check the reason s/he visited the facility, e.g. bulletin board announcement of special materials available, teacher assignment, guest speaker, etc.; analysis of reasons could tell you what motivates students to use the facility. Suppose you need to know the adequacy of the facility's resources. To determine this you can maintain a record of requests and determine whether there is need for items not held in the facility, and whether a particular item or class of items is in heavy demand. This information will help you decide which additional

resources are needed. It may also affect your decision to alter the facility's size or location if you are going to have a major increase in the number of holdings. But be sure to relate purchases to evidenced student outcomes.

Step 3: Develop the Evaluation Design

The procedures you use to collect data necessary to answer your questions and to analyze those data will vary according to whether you are dealing with outcome or process questions. Process questions to determine how well the facility is functioning are relatively straightforward. For example, if you want to assess the degree to which users find what they need when they come to the facility, you could ask students, teachers, and community representatives to rate their experience with facility personnel on a scale from 1 to 5 and to suggest competencies or roles they find lacking. To analyze these data you would compute the mean rating of staff performance and compile a summary of suggestions for staff improvement.

Outcome questions require a different approach. After collecting data on the performance of students (or teachers) who have used the facility, you will need to compare those data with a standard of some sort; this will strengthen your ability to attribute any differences that appear to facility use. However, there may be plausible alternative explanations. The plausibility of these alternative explanations should be carefully examined before evaluation results are attributed to facility use. The key is to use common sense in preparing a case that something happened that was the result of facility use by students and/or teachers. First you need to establish a baseline for how teachers and students would perform if they didn't use the facility. There are a number of ways to do this. For example, if your school has administered a standardized test of career development during the last two or three years, you can examine trend data to compare student scores on decision-making skills before the facility was established with student performance after using the facility. This is a descriptive design. Differences in student outcomes can be analyzed by computing a correlation to determine whether career maturity and facility use covary.

If trend data are not available, you can use national norms as a comparison standard. For example, you can administer a standardized career development test on a pre/post basis to a random sample of students in your school who use the facility. If you find that your group of students has a higher percentile status after facility use than before, you can possibly conclude that the facility was effective. [However, this inference could be distorted; see the Ideabook (Tallmadge, 1977) for errors to be avoided when using normative data.]

Step 4: Identify/Develop Instruments to Collect the Data

In Step 1 you specified the questions you want to answer in your evaluation. Now it is time to select/construct measures that will provide those answers. The measures you use must meet two criteria: they must address your specific information needs and they must be appropriate for the population(s) you want to assess. As in Step 3, the instruments you use will vary in relation to outcome and process questions. To answer process questions, you may develop student, professional staff, and community evaluation forms. On these forms respondents can answer yes-no kinds of questions concerning facility functioning and can assess services in terms of degree of helpfulness.

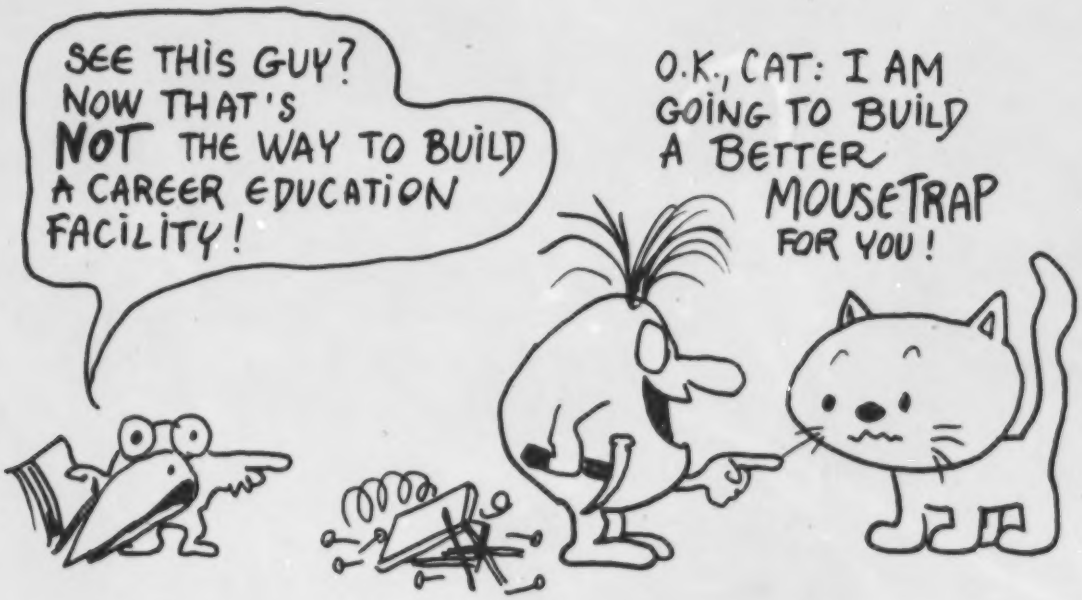
Identifying instruments to assess student outcomes can be more difficult. There has been a general lack of information about career education measures. Ohio State's Center for Vocational Education has developed a handbook that provides descriptions of an assortment of measurement instruments currently being used to assess career education objectives (McCaslin, Gross, and Walker, 1977). Inspecting this handbook can give you information about standardized instruments that may match your information needs.

If you don't find any instruments containing the types of indicators you specified as needed to furnish evidence of attainment of the facility's goals in terms of student outcomes, you can develop your own. Locally developed instruments should be constructed according to accepted guidelines (see McCaslin and Walker, 1978). First establish the real purpose of the instrument by asking whether you need instruments designed to: (1) assess students' specific knowledge; (2) assess students' ability to make decisions; (3) determine students' ability to perform tasks

from which you can infer that certain skills have been acquired; and (4) assess some combination of the previously mentioned attributes. Second, identify the various uses that can be made of the instrument. The primary use probably will be to assess facility effectiveness. However, the instrument might also be used to diagnose student weaknesses; activities can then be prescribed to correct identified deficiencies. Third, determine whether your instrument seems to focus on trivial matters; items that ask simply for recall of certain facts will not provide as much information about the effectiveness of the facility as those that test for understanding or ability to apply specific knowledge.

After your instrument has been developed, determine its reliability and validity. By reliability we mean the instrument's ability to provide similar results each time it is used. There are three key causes of unreliable instruments: (1) inadequate directions; (2) ambiguous items; and (3) poor scoring procedures. By validity we mean the instrument's ability to measure what it is supposed to measure. Pilot test the instrument (i.e., administer it to a small number of students, perhaps 15) to see how various items function. McCaslin and Walker (1978) provide practical suggestions for revising an instrument to improve its reliability and validity.

Suppose one of your facility's goals is to increase students' ability to identify options for recurrent career-related training in several job areas. You could develop a questionnaire asking students where they would go for training after having been in the work force for five years. Or you could develop a multiple-choice instrument on which students indicate the most appropriate source for different kinds of training. After the instrument has been pilot tested and revised, administer it to student facility users each year, first to establish a baseline and then to analyze trend data.


SEE THIS GUY? NOW THAT'S NOT THE WAY TO BUILD A CAREER EDUCATION FACILITY!
O.K., CAT: I AM GOING TO BUILD A BETTER MOUSETRAP FOR YOU!

Step 5: Set Up the Logistics for Data Tabulation and Analysis

If you have completed Steps 1-4 your evaluation has gotten off to a good start. You have specified the questions you want to answer and have identified the instruments you will use to obtain needed information. Now you need to set up a system that will ensure unbiased data collection and accurate data tabulation and analysis. This is an important step, for if your evaluation results in unusable data, your carefully specified questions and resulting decisions will be in limbo.

Break down the data collection tasks and assign each component to a particular person, along with setting a task completion date. This task breakdown should be circulated to all pertinent personnel so that everyone knows what should be occurring at a particular time. Be flexible. If problems arise, pull the team together and make decisions not only about the scheduled tasks, but also about future tasks that might be affected by any change or delay. Remember that the focus of your evaluation is to increase the effectiveness of your facility in terms of student outcomes, not to prove that it is or isn't working.

If you are collecting trend data, you need to identify information sources and to develop a checklist to ensure comprehensive searching of sources and recording of findings. If your design includes teacher-administered tests, you should institutionalize procedures to reduce the

possibility of attributing test results to unintentional teacher bias in test administration and scoring. You can introduce standardized procedures for test administration during staff development sessions. You can provide followup assistance and occasionally spot-check test administration and scoring.

Finally, keep in mind your data analysis strategy when you are reducing the data. We know of cases in which average student scores were computed and the raw scores discarded without determining the standard deviations, making it impossible to perform a t-test.

Step 6: Use the Evaluation Results

Now you have the ingredients for determining the strengths and weaknesses of the facility - the student outcomes it facilitates and those that are unaffected by the facility. You have information about how and why the facility was successful or unsuccessful in terms of each desired student outcome. You may have identified some side effects, such as reduced class-cutting or absenteeism, or increased diligence in preparation of written reports.

Report your findings in a form that is usable by decision makers. Remember that negative results are just as important as positive ones in terms of helping you sharpen the effectiveness of your facility. Also remember that there was no point in conducting the evaluation if you do not use the results.

Sample Executive Summary

The following pages present an executive summary of an evaluation of a hypothetically operating career education facility. This will help you understand the kind of information you can derive from such an evaluation, and will suggest uses of such information.

PROJECT CUCE (Combat Unrealistic Career Expectations)

Executive Summary

A longitudinal study of students in grades 9-12 found that by the time those students had reached age 30, their major criticism of their high school experience was its failure to help them develop long-range educational goals or to plan for a career. Young adults viewed this deficiency in career guidance as the school's greatest failure. Unrealistic career expectations resulted from students lacking an understanding both of the education needed for particular jobs and of their own abilities and interests. After reading the results of this longitudinal study in a newsletter published by a federal-level education agency, staff at Anthony Wayne High School decided to investigate whether their students had the same information that the national study had revealed. A formal needs assessment disclosed that students had limited knowledge of the qualifications needed to obtain a variety of occupations as well as an inability to incorporate their self-concepts in planning for their own career development. Therefore, the school opened a career information center with two major goals:

- to provide information on specific careers and the preparation they require; and
- to equip students with a degree of self-understanding sufficient for making sound career decisions.

Method

The career information center was housed in the school library and was designed to integrate career planning concepts into the school curriculum through a library-based program of career information services and self-awareness activities. A center coordinator was hired on a full-time basis; students assisted in maintaining an inventory of resources and in checking materials in and out of the center. The center coordinator conducted inservice workshops to introduce the faculty to the center's services; he provided the teachers with subject bibliographies of the center's holdings to make them aware of the materials available to help them integrate career education into their teaching. After obtaining teacher reactions to the bibliographies, the coordinator increased the number of audiovisual resources and returnable books and decreased the number of hand-outs available through the center. The coordinator developed an independent study mini-course that first exposed students to materials that increased their awareness of their own interests and abilities and then referred them to materials that illustrated the importance of selecting jobs whose properties dovetail with student characteristics.

The target population consisted of students attending a suburban high school. Twenty percent of the students came from high-income families, 58% were from middle-income families, and 22% came from low-income families. Seventy percent of the student participants were white, with the remaining 30% composed of Native American, Black, Asian American, and Latin American participants.

The district evaluator worked with the center coordinator from the beginning of the project. She specified tasks concerned with general logistical aspects of the evaluation design as well as more discrete steps in collecting and analyzing data. The tasks were broken down, and each component was assigned to a particular person, along with a task completion date. Since the design included teacher-administered tests, the evaluator institutionalized procedures to reduce the possibility of attributing test results to teacher bias in test administration and scoring. Standardized procedures for test administration were introduced during inservice training. Follow-up assistance was provided to teachers, and test administration was spot-checked. Since teachers scored tests and recorded results for their own students, the evaluator checked the scoring and randomly inspected student answers for indications of possible bias in test administration.

The evaluation design included administration of two tests--a commercial, standardized test in which student scores were compared with national norms, and a locally-developed test, in which trend data were analyzed. The project-developed instrument consisted of 40 multiple-choice items and was designed to measure knowledge of selected occupations and the training/education required to enter them. During the first project year the locally-developed instrument was pilot tested, revised, field tested, and revised again. Examples of a wide variety of careers and occupations were included in the items. The following steps were taken to eliminate stereotypes:

- Both masculine and feminine pronouns were used in the items, neutral terms were used in describing various careers, and both males and females were represented in nontraditional roles and occupations.
- Names representing various ethnic groups were used in the instrument.
- Items reflecting a variety of life styles were included (e.g., single- and two-parent families, extended families, and suburban/urban/rural settings).

Test administration during the first year furnished baseline information. In subsequent years, trend data were analyzed. As shown below, students in the target year (the second year of center operation) achieved significantly larger scores than those achieved by students in the baseline year.

	Prior Year	Target Year	p
N	101	134	
$\bar{X}$	28.5	30.6	< .01
SD	13.5	14.8	

The commercial test utilized was the Competence Test of the Career Maturity Inventory (CMI). This 100-item test is concerned with knowledge about occupations and career selection decisions. The Competence Test consists of five separate sections: (1) self-appraisal, (2) occupational information, (3) goal selecting, (4) planning, and (5) problem solving. The test was administered on a pre-posttest basis. Student scores were analyzed in relation to national norms. Raw scores of individual students were used to compute mean percentile levels. As shown below, growth in relation to national norms was statistically significant.

	September 1977		May 1978		Gain in	
N	Raw	Percentile	Raw	Percentile	SD Units	t
134	29	8	48	22	.63	7.30*

*Significant at .001 level, two-tailed test.

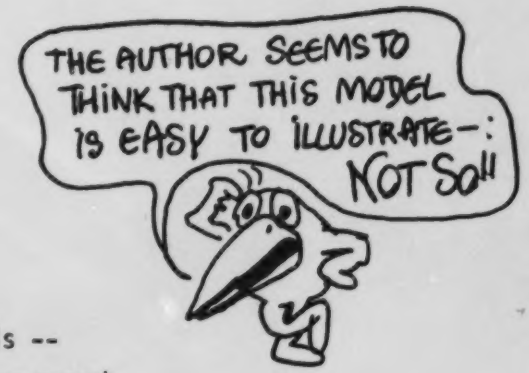
Follow-up Implications

Both measures of impact provide clear evidence that Project CUCE has been successful in increasing students' awareness of the preparation required for a variety of specific careers and in equipping students with the ability to make sound career decisions on the basis of individual interests and competencies. The locally-developed instrument has high face validity in that it is closely related to the first project goal. In addition to the growth demonstrated in comparison with a local standard, the commercially available test demonstrated that students made a significant gain in both goal areas in comparison with a national norm. Since the project was successful in achieving both goals, it appears to be a feasible and effective program--one that should be continued.

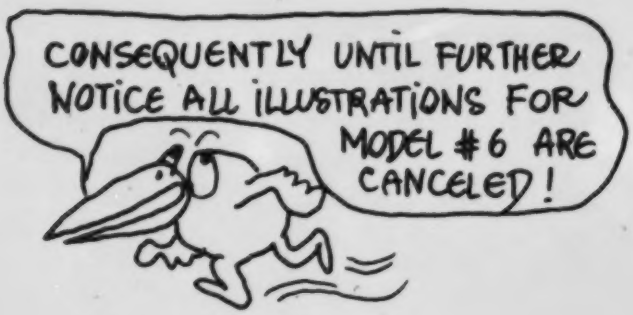
EVALUATION OF INDIRECT INTERVENTIONS

Model # 3

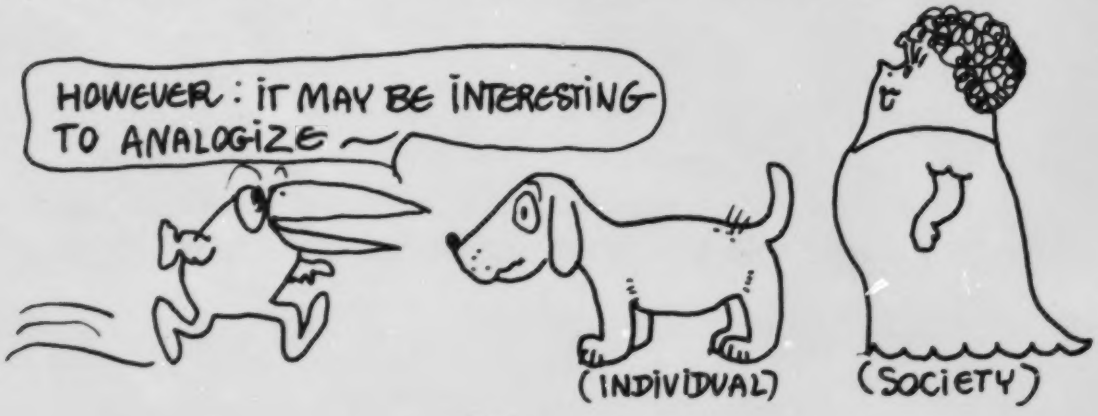
Anita M. Mitchell



Late twentieth-century education in America appears to be trying to mold individuals to cope with constraints imposed by societal structure. Most career education activities are designed to help the individual acquire specific knowledge/competencies in the areas of self understanding, knowledge and understanding of the world of work, and decision-making skills -- preparing the individual for society. In our search for exemplary evaluated career education activities we found <u>none</u> specifically designed to prepare society for the individual.



Societal institutions generally are assumed to be unchangeable (even though we may grumble about the barriers they present to our clients), so we encourage young people to be aware of these obstacles and to modify their own needs/interests/talents to match societal needs and opportunities. Yet social systems research presents an array of evidence that programs focused on changing the system have potential for positive impact on students, with far greater efficiency than those focused on changing individuals.



You, the career educator, know what is frustrating young students and graduates of your school as they seek to pursue educational and occupational choices. You've decided that your program will not begin by working directly with students; rather it will work first with those institutions and policies that stand in the way of students' career development. You have decided to make positive effort to facilitate changes in the contexts and institutions that affect students' education, training, and employment opportunities and options. Instead of helping students directly, you will be helping them through indirect interventions in career education.

Indirect interventions include the following types of activities:

- Activities designed to affect local board policy or state legislation related to career education.
- Activities designed to encourage publishers to remove sex stereotyping and other bias factors from educational materials and instruments.
- Activities designed to increase local job training options.
- Activities designed to make available to students full and factual, rather than biased or distorted, occupational information.
- Activities designed to expose students to a wide range of positive role models.
- Activities designed to increase parents' awareness of, and ability to counteract, those subtle but pervasive societal influences that tend to limit occupational options.
- Activities designed to provide open-entry, open-exit education.
- Activities designed to inform/affect local advisory committees.
- Activities designed to affect labor union policies.
- Activities designed to increase public awareness of career education needs.
- Activities designed to motivate teacher/counselor/administrator training institutions to include career education in their training programs.

This list is not exhaustive; rather, it is representative of the many activities schools might undertake to open options and to remove societal barriers that limit young people's educational and occupational aspirations and their educational and occupational choices.

Indirect intervention activities are clearly implied in the Career Education Incentive Act which states that a major purpose of education is "to prepare every individual for a career suitable to that individual's preference" and that "educational agencies and institutions...should make every effort to fulfill that purpose." (Sec. 2, (1) and (4).) An indirect intervention that would contribute to this educational purpose would be to encourage the local Chamber of Commerce to develop and maintain a directory of businesses and industries in the local and surrounding communities, including brief descriptions of all the various jobs held by employees of each of these businesses or industries. Such a directory should refine students' perceptions of the kinds of business/industries that employ workers in the areas of the students' preferences, as well as whether the need for workers in those areas is broad or limited.

Sec. 3 stresses "collaborative arrangements with the appropriate agencies and organizations." The same example as above might be used to illustrate indirect interventions related to this section of the Act; the intervention would be made collaborative by having counselors furnish the Chamber of Commerce with a list of students' occupational preferences, and encouraging them to feature these preferences in their directories. Sec. 10 states that funds reserved under section 5(a)(2)(A) of this act may be used to develop career education projects "designed to elminate bias and stereotyping on account of race, sex, age, economic status, or handicap." An excellent example of an indirect intervention in this area -- one that should be pursued by every school -- is a systematic assessment of all career education materials made available to students in the school, with the focus both on eliminating those found to have bias or stereotyping, and on encouraging the publishers of those documents to revise them for future use. Since Congress considered these directives so important to the career development, career opportunities and career success of the nation's youth, it is incumbent upon us that we move swiftly to implement the Act, with or without funding.

In his Primer of Career Education, Dr. Kenneth B. Hoyt,[1] Director of the Office of Career Education, lists nine learner goals as foci for career education activities. Particularly related to indirect interventions are goal 6 (aware of means available to them for continuing and recurrent education) and goal 9 (aware of means available to themselves for changing career options -- of societal and personal constraints impinging on career alternatives). Goal 6 clearly suggests such indirect interventions as bulletin board displays prepared by representatives of a variety of postsecondary education and training options, or appealing to these representations to present informational and motivational vignettes on local TV. Goal 9, while it assumes a direct intervention to educate students to be aware of constraints and of ways to remove these constraints, also requires that the school educate the community so that they are aware of the effects of their actions and images on the career options of students. A vigorous campaign may be needed to make the community more receptive to students' efforts to move beyond the real or imagined constraints with which they are confronted.

Obviously, evaluation of the effectiveness of indirect interventions will differ from evaluation of direct interventions with students. Having identified the institutional change you hope to effect, and having stated a viable rationale for your focus, you will want to determine exactly which activity or activities would be most likely to accomplish your objective. Also, your rationale will have explained how the institutional change will positively affect students. Evaluation of the effectiveness of your strategy is relatively straightforward -- the change has been totally accomplished, partially accomplished, or there is no change. A critical question, however, is whether the institutional change made a difference to students: did they change their educational or occupational aspirations as a result of the successful indirect intervention? Did they move into the newly-opened educational and occupational options? Your evaluation, then, will focus on two results: (1) whether the change occurred, and (2) how student behavior was affected by that change.

In indirect intervention, it is quite possible that the institutional change may be accomplished within a single school year, but it is unlikely

---

[1] Hoyt, Kenneth B. Primer of Career Education. Washington, DC: Office of Education, Office of Career Education, U.S. Department of Health, Education, and Welfare, 1977.

that the results in terms of effects on students will be known for at least another year. Therefore we are talking about a multi-stage or at least a two-stage evaluation, and one that probably will require data collection and analysis over a period of two or more years. In your situation you may find it adequate to evaluate the first stage only: did the institutional change occur? Or you may want to move on to the second stage: did this change affect student aspirations and choices? After the indirect intervention was accomplished, did you need some type of direct intervention with students to encourage their taking advantage of the newly opened options or was awareness of the institutional change sufficient? The third stage of your evaluation would be: did the institutional change and subsequent student participation in new options result in the positive student behaviors you envisoned?

The three stages described above are discrete and you may undertake any one of them independently. You do not have to decide in advance whether you will implement evaluation for one stage or two or three; but you will need to conceptualize all three evaluations so you will have a consistent and reliable model when and if you do decide to implement successive stages, and so needed date will have been collected.

In this type of activity, since in most cases the intervention requires the assistance of persons outside the school, there is a need to collaborate with community individuals and/or agencies from planning through implementation and evaluation. Without such collaborative effort it may be impossible to gather accurate data which will give evidence that the institutional change did make a difference to students. Without such data, the fact of your having accomplished your primary objective -- of effecting an institutional change -- will be of little value; certainly it will not encourage others to try to effect similar social change. The purposes of evaluating the effects of indirect interventions are: (1) to show the extent to which social institutions can be changed through local effort, (2) to identify effective vs. ineffective strategies, (3) to determine whether or not the effort was justifed in terms of the outcomes, and (4) to prove that institutional change has a positive effect on students.

Decisions

There is no point in conducting an evaluation if you are not going to use the results. Evaluation data furnish information for decision makers. The

information collected and analyzed will be that information needed to make decisions. Who will be making decisions for which evaluation data from your indirect interventions will be useful? The school board? The superintendent? The principal? Teachers? Community organizations? Students? Identify the decision makers first, and let them identify the decisions they need to make. A variety of decisions might be made based on the evaluation results. Decisions could include:

- Whether it would be feasible to plan other indirect interventions.
- Which agencies to include in program efforts.
- Whether there are more cost efficient ways to achieve the same results.
- How to multiply the effects of the activity.
- How to modify future efforts.

Questions to be Answered by Evaluation

The questions to be answered by the evaluation will relate specifically to the problems and needs you identified when you decided to pursue this particular activity. Questions will fall into two categories. Types of questions you may want to ask in each category are presented as examples only; no one of these questions is automatically appropriate to ask for your activity.

Process questions

- What happened?
- How did it differ from what was supposed to happen?
- What were the reasons for discrepancies?
- Were the plans implemented as intended?
- Were there unanticipated problems in communication between school and other agencies?
- What were the keys to success? Consider processes such as communication, assistance, product delivery, conflicts of values, and actors such as agencies, project staff members, community members, etc.
- Which agencies were most helpful?
- Did any agencies fail to cooperate?
- Was it necessary to modify plans in order to reach the objective?
- Was the project team (local staff and community extensions) qualified to accomplish the tasks assigned?
- What kinds of discrepancies in implementation were found?
- Was the objective found to be unrealistic or unattainable?
- Was compromise necessary?
- Was the time schedule realistic?
- Were deviations from the time schedule due to planning error or to human caprice that could not have been anticipated?

Outcome questions

This group of questions is divided into two sections: (1) outcomes in terms of the institutional changes intended, and (2) outcomes in terms of effects on students.

Questions in the first category could be:

- Did the change occur (e.g., was the law enacted, or the policy changed, or the printed material modified, etc.)?
- Is the nature of the change such as to reasonably assure that it meets the need for which it was intended?
- What were significant side effects, if any?

- Was any protocol developed for effecting similar changes in the future?

Questions in the second category could be:

- How many students benefited from the change?
- Does the change benefit various subpopulations differentially?
- How is student benefit evidenced?
- Does the benefit to students justify the cost of the activity?
- Is one approach more effective than another?
- Do the benefits to students warrant pursuit of additional institutional changes.

Advantages and Disadvantages

The major advantage of using this evaluation model is that it provides for documentation of evidence that the institutional change you intended was or was not accomplished (or, if partially accomplished, the degree of accomplishment and the effect of partial accomplishment on intended student outcomes). Another advantage is that it provides for presenting evidence of the impact of your activity on students.

A disadvantage is that, particularly at stages 2 and 3, you will need to depend upon some non-school persons to collect and report data. Since your program is intended to affect some institutional change (in your own school or in another community or state agency), determining the impact of this change on students requires your seeking information about students' experiences as a result of that change. Some of this information may be obtained from employers or employment agencies, and some may be obtained directly from the students. Insuring the accuracy of this information can present some rather serious problems. For this reason it is important to identify your data sources before you implement your activity, and to orient key source persons to the need for promptness and accuracy in data collection. Also, it will be important for you to orient the evaluation sample of students to the fact that they are to be involved in a study lasting at least two years, and to secure their assurance of cooperation.

Steps in Implementing This Evaluation Model

If your program focuses on one or more indirect interventions and you have decided to use this evaluation model, you will need to complete a number of planning and implementing steps. Do not allow the number of steps to discourage you; actually, many of them require very little time; many of them are things you do anyway when you are planning your program. The difference is that you will write them down so you can check as you go along to make sure there is no mismatch among or between components of your evaluation. Unless you commit your evaluation plan to paper, you are likely to miss some logical fallacies that might be built in. You need to be able to conceptualize and communicate a complete understanding of your evaluation system, with particular attention to a rigorous fit among your data needs, data management and data analysis.

Although this is not a program planning model, but an evaluation model, the two cannot be totally separated, so you will find that there are evaluation tasks at each step of the planning process. In other words, a good evaluation system is a necessary component of a good management system, with evaluation concerns affecting each program planning decision, and evaluation data affecting each step of implementation.

As you move through the steps you will find that there are logical connections between and among them. Decisions made at one step affect decisions at later steps. Decisions at later steps must relate directly to decisions made at earlier steps. For instance, when you get to Step 6, where you are selecting the instruments to be used for data collection, you will refer to Step 4 -- the point at which you defined your evaluation <u>needs</u>. If there is not a good fit between these two steps, your evaluation will not serve the purposes (Step 2) or provide data for decisions.

As you go through this section of the evaluation model you will find the following steps:

Step 1 - What to Evaluate (the total program? One component?)

Step 2 - Purposes (Why bother? Will you use the results?)

Step 3 - Identification of Evaluation Resources (Do you have access to persons skilled in evaluation? Is there enough time? Money? Equipment? Materials?)

Step 4 - Evaluation Needs (What information do you need to serve the purposes you specified in Step 2?)

Step 5 - Evaluation Design (What comparison standard will you use so you'll know that changes in the target population were due to the treatment?)

Step 6 - Instrumentation (What tests, observation techniques, questionnaires, demonstration tasks will provide the information needs you specified in Step 4?)

Step 7 - Data Collection and Management (How can you make sure the data collected are accurate and the data reduction procedures have preserved that accuracy?)

Step 8 - Monitoring System (How can you make sure the activities are being implemented as planned?)

Step 9 - Reporting (How are you going to get evaluation results to decision makers in a form that they will understand and use?

Step 1: Determine What You Are Going to Evaluate

In order to determine what it is you are going to evaluate you must first describe what it is you want to accomplish at each stage. For Stage 1, you may ask what is the activity all about? What is it expected to accomplish? What institutional factors are expected to be changed? For Stage 2, how many students are expected to be affected by your indirect intervention? How do you expect that they will be affected? For Stage 3, you may ask: With the institutional change accomplished and student access facilitated, what is expected to be the impact on students?

Be precise in stating feasible, important objectives, e.g., Stage 1: "District policy will be revised to allow graduation credit for exploratory work experiences in the community"; or "There will be an increase of 200% in the numbers of local businesses offering work experience opportunities for handicapped students"; or, "Administrative guidelines for State proficiency examinations will provide for differential standards for bilingual students."

Stage 2 objectives might include "All students will participate in an exploratory work experience related to their occupational interests" or "All handicapped students will participate in work experience programs in the community"' or "Bilingual students of average ability, after at least two years in _____ School, will score at pre-specified acceptable levels for their cultural group on the State proficiency examinations."

Stage 3 objectives will focus on the impact on students, e.g., "Students participating in exploratory work experiences in the community selected elec-

tive courses related to tentative occupational goals"; or "Handicapped students participating in work experience programs in the community showed significant gains in attitudes toward self, attitudes toward work, realistic but ambitious occupational aspirations and specific occupational competencies"; or "Bilingual students scoring at acceptable levels on the State proficiency examinations were significantly more successful in finding jobs than were comparable students before State proficiency standards were changed."

As you can see from the above objectives, you need to identify the target populations: Stage 1 - the institution (school, community businesses, state department of education, legislature, etc.); Stage 2 - the students who are expected to benefit from the institutional change, and Stage 3 - the stucents who took action related to the change.

Step 2: Purposes of Evaluation

The next step in your evaluation is to define the purposes of your evaluation. The purposes of evaluation are closely tied to purposes of the intervention. There are general purposes such as furnishing information to decision makers, giving participants some feelings of accomplishment, communicating accomplishments to the public. More important is to state the rationale for what the activity plans to accomplish. In other words, why are you doing it? What are the problems you expected to solve by this activity? What are the needs that will be met by the results of this activity? For instance, using some of the examples in Step 1, we could say that the problem is that the district policy includes so many academic requirements that the students can't fit work experience into their schedules. The need is for more flexibility in requirements, so students can explore occupational fields. Or, the problem is that two few local businesses will accept handicapped students in work experience programs; the need is for these students to have the same kinds of experiences as nonhandicapped students. Or the problem is that bilingual students are denied diplomas because of state-imposed standards that discriminate against them; the need is for differential standards.

Problems and needs requiring indirect interventions may be identified through a comprehensive needs assessment. It is more likely, however, that the problem and need will be manifested through examination of students' choices of education, training, and/or occupations. Counselors frequently are aware of badly skewed distributions of student choices in each of these

STAY AWAY FROM ME, YOU DIRTY DOG! DOGS ARE COLORBLIND; DOGS HAVE FLEAS; DOGS ARE NOISY; DOGS SHED HAIR; DOGS BITE; AND BY THE WAY: ONLY FEMALES ARE FAITHFUL TO THEIR MASTERS !!
THIS CALLS FOR INDIRECT INTERVENTION!

areas; analysis of the distribution will usually reveal societal/institutional conditions or policies which might be changed through the efforts of school personnel.

Your statement of purposes, then, will be related to your needs assessment information, and will give a rationale for choosing this particular activity to solve the problem(s) and meet the need(s). Precision in this statement helps you focus on precisely what variables are to be measured.

Step 3: Identification of Evaluation Resources

Next you will need to identify resources and constraints. Human, technological, material, and financial resources need to be inventoried. Are there persons employed by your school or school district who have the knowledge, experience, time, and interest needed to perform such tasks as selecting/modifying/developing instruments? Developing data management systems? Selecting and performing appropriate data analysis techniques? Writing evaluation reports? etc. Do they have the time and interest to perform these tasks? If not on the staff, where can such experts be found, and how much of their time will you need? Who are the persons in the community who can provide you with data you need, and how can you secure their cooperation? How much will it cost to secure such help? Is that amount of money available to spend on these tasks?

Technological resources include equipment and services, e.g., what is the cost and accessibility of such resources as computers and computer processing? Material resources include supplies, and financial resources include the money needed to support the evaluation staff, plus their technological and material needs. For instance, if you are trying to increase the number of businesses accepting handicapped students for work experience, you probably will need a cadre of contact persons to make personal calls on business executives to insure their cooperation in meeting your objectives and in supplying such information. Secretarial, mail, and telephone costs should also be included in your inventory of resources.

An analysis of confirmed resources (Not "I know Charles Smith knows how to develop instruments" but "I know Charles Smith knows how to develop instruments and he has agreed to work 10 days in March on this evaluation") will help you select a feasible evaluation design: one for which there are adequate human, technological, material, and financial resources. When looking at financial resources needed, be sure to consider costs of such items as the following:

CAREER EDUCATION PROGRAM EVALUATION BUDGET

LINE ITEM	
PERSONNEL COSTS	$_____
A SALARIES & LEAVE	$_____
B FRINGE BENEFITS	$_____
OVERHEAD	$_____
TRAVEL	$_____
A LOCAL MILEAGE	$_____
B CONSULTANT PER DIEM	$_____
C OTHER	
OUTSIDE SERVICES	$_____
A PROFESSIONAL CONSULTANTS	$_____
B SUBCONTRACT	$_____
C OTHER	$_____
OTHER DIRECT COSTS	$_____
A MATERIALS	$_____
B EQUIPMENT	$_____
C POSTAGE	$_____
D XEROX	$_____
E TELEPHONE	$_____
F EXPENDABLE SUPPLIES	$_____
G OTHER	$_____
TOTALS	$_____

Step 4: Evaluation Needs

Now you are ready to determine your evaluation needs. First identify the questions to be asked -- including those to be answered while this program is underway to improve the likelihood of significant impact. Steps 1 through 3 will help you formulate questions to be asked. Make sure teachers, students, parents, and community representatives have an opportunity to suggest questions they want answered. Questions for Stage 1 might include "What is the most effective way to motivate business executives to provide

work experiences for handicapped students?" "Is one type of business more open to handicapped students than other types of business?" For Stage 2, "Do handicapped students seek placement in newly available work stations in community business?" "Are additional strategies needed to facilitate their participation?" For Stage 3, "Are handicapped students able to function in the types of work experiences being provided?" "Are certain kinds of experiences ill-advised for certain types of handicapped students?" "Does the work experience make a difference in the handicapped students' self concept and occupational aspirations?" "Do handicapped students participating in community work experiences develop work competencies they could not develop in the school setting?"

This is probably THE most important step, because it brings into focus the purposes, the institutional change sought, the target population, the

treatment strategy, the probability of effectiveness, and reasonable standards by which to judge effectiveness. Without this step you might gather data that really do not illuminate the central issues. Write all the questions you would LIKE to have answered at each stage; then reduce the number of questions, retaining only those for which answers are needed in order to make decisions. For each question, ask "Why do we need to know this? How will we use the information? What important decision(s) will it help us make?" If the answer does not contribute to a decision that will actually be made, the question should not be asked. Rank your questions in order of importance, to facilitate development of indicators from which you can infer that the objective has been met. This step will help you identify your evaluation data requirements. It may help in this step first to state decisions, then to identify hypothetical statements that would serve as rationales for making the decision one way or another, and finally to state information needs as needs to verify those hypothetical statements or rationales.

Information needs will determine most appropriate data sources. For example, for Stage 1, in order to decide whether or not to continue to encourage local businesses to provide work experiences for handicapped students after completion of the current program, you might want answers to such questions as "Has there been a significant increase in numbers of businesses offering work experience to handicapped students?" "Would these businesses have made these opportunities available without our intervention?" "Are the experiences offered representative of the kinds of experiences handicapped students need to develop self confidence, to focus on occupational goals, to develop job seeking and job getting skills?" Indicators would include actual counts of verified work experience stations after the intervention as compared with before the intervention; self reports from employers indicating change in their willingness to accept handicapped students for work experience before and after the intervention; analysis of the types of experiences available after the intervention. Data sources would include past records and current employer reports; job analysis data and student need data.

Stage 2 decisions might be whether or not to use direct interventions with students in order to stimulate them to try newly available work stations. Questions might be, "Did handicapped students take advantage of the

work experience openings voluntarily, after being made aware of the openings?" "What, if any, procedures were needed to facilitate students' taking advantage of the openings?" "Did one procedure work better than another?" "Are there unanticipated problems, such as transportation to the work experience site?" "Are the openings consistent with the interests and aptitudes of the handicapped students?" Indicators would include an actual count of the number of available spots that were filled by handicapped students; documentation of procedures needed to inform/persuade/coax/coerce handicapped students to take advantage of the openings; documentation of which procedures worked and which did not; self-report information from employers and students concerning unanticipated problems; analysis of openings as they relate to interests and aptitudes of the students. Data sources would be past and current records; reports of persons implementing facilitation procedures; student interest and aptitude test data; job analysis data; and interviews with employers and students.

Stage 3 decisions might include whether to install in-school programs for helping handicapped students develop a specific set of competencies, and which competencies best lend themselves to in-school instruction. Stage 3 questions could include, "Have handicapped students participating in work experiences in the community developed career competencies that could not have been developed in the school?" Indicators would include evidence of competencies developed and evidence that those competencies could not have been developed within the school setting (with current resources). Data sources would be the students themselves (pre- and posttesting of competencies), and school staff (analysis of resources available for developing these competencies within the school).

Step 5: Evaluation Design

Now it is time to develop the evaluation design: the way comparisons will be made between post-intervention data and a comparison standard. The limits of length and complexity in this document preclude our displaying all the possible types of comparisons you might make. However, it is important for you to be aware that there are many kinds of questions you might ask, and that the kinds of questions will determine the kinds of analyses which will be appropriate. The following questions illustrate the various possible approaches:

1. "Did the intervention significantly increase the number of work experience stations available to handicapped students?"

2. "Did the intervention double the number of work experience stations available to handicapped students?"

3. "What does it cost to create one additional work experience station for handicapped students?"

4. "Which of two or more approaches is the most effective in creating work experience stations for handicapped students?"

Question 1 requires comparison between treatment and no-treatment effects. Question 2 requires comparison between treatment effects and an absolute standard (doubling the number of stations). Question 3 requires computation of the actual cost per station. Question 4 requires comparison of one treatment with another.

When to derive no-treatment expectations. Our illustrations here will be limited. If you are asking Question 1-type questions, an essential task will be to derive no-treatment expectations. In other words, for Stage 1, how many businesses would accept handicapped students for work experience if you did NOTHING; for Stage 2, how many students would participate in work experiences in the community if you did NOTHING; for Stage 3, what career competencies would handicapped students have at the end of the program if you did NOTHING? You will need to make some credible estimate of conditions without intervention. This estimate may be made in a number of ways. For this model, for Stages 1 and 2, the best way to establish no-treatment expectations is to examine trend data. These data are probably available in your records; if not, you may need to survey local businesses. You will need to know how many (and what percentage of) handicapped students were placed in work experience programs in local businesses during each of the last three (or more) years, how many businesses provided these work experiences for handicapped students, and what types of experiences were provided. These trend data, plus an assessment of other efforts being made that would produce similar effects, will allow you to derive your no-treatment expectations. For Stage 3, trend data may not be available on career competencies of handicapped students over the past three years, so you will need to derive no-treatment expectations in this area from a different set of information. It might be feasible to find a similar population of handica-pped students in another school in your district or a neighboring district, and to establish a baseline by conducting a competency analysis on both the treatment and no-treatment students. The two groups would be compared at

the end of the intervention, to determine the effect of the intervention.

Once you have established your procedures for estimating no-treatment expectations and have decided on the kind of data you will be collecting and comparing, you will need to determine the type of analysis to be used to interpret the evaluation data.

For Stage 1 of our example, one of our questions was, "Did the indirect intervention significantly increase the number of work experience stations for handicapped students?" This does not require a sophisticated evaluation design. The design is descriptive rather than experimental, and all that is required is to collect and compare data after project implementation with historical data before project implementation. A simple t-test will tell you whether the changes are statistically significant.

For Stage 2, one question was "To what extent did students take advantage of the new options?" This requires collection and analysis of data on: (a) numbers and percentages of students in work experience programs as compared with previous years; (b) length of each experience as compared with previous years; (c) types of experiences as compared with previous years; and (c) types of experiences as compared with previous years. (b) and (c) data are harder to collect, as there is so much opportunity for unintended distortion. This will be discussed further under Instrumentation. As in Stage 1, the design is descriptive, and analysis of the statistical significance of differences before and after your intervention is best accomplished by applying a t-test.

Stage 3 focuses on career related attitudes and competencies developed by handicapped students participating in work experiences in community businesses. Here a pre- and posttest design is of little value if there is not a control group with which to compare the treatment group. Handicapped students not participating in the program could be compared with those participating only if the two groups were comparable on such variables as sex, grade level, ability level, degree and type of handicap, etc. A treatment/control posttest-only design might be preferred, with control students identified from another school or school district. Again, statistical significance of comparative data can best be determined by applying a t- test.

Using an absolute standard. If you are asking Question 2-type questions (e.g., "Did the intervention double the number of work stations...?"), comparing the effects of your program with an absolute standard, your approach will be slightly different. Obviously, you will not need any statistical analysis to determine the answer to your questions. A simple arithmetic computation will suffice. However, if you do not achieve your goal of doubling the work experience stations, you will need to fall back to a statistical analysis of the differences between pre-treatment and post-treatment data to determine whether the differences can be attributed to the intervention. Here again, a t-test will be appropriate.

Determining costs. If you are addressing a Question 3-type question (e.g., "What does it cost to create an additional work station...?"), you will need simple straightforward reporting of costs, and your analysis will be a matter of judgment as to the cost benefit of your efforts. Such cost benefit should be determined in relation to a predetermined standard, e.g., "We believe it is worth $1000 to achieve each new work station for a handicapped student." Determination of effectiveness of the program is achieved by computing actual costs of increasing work stations, with the amount you were willing to spend.

Comparing treatments. If you are addressing a Question 4-type question (e.g., "Which of two or more approaches is most effective...?"), it is assumed that you are comparing two or more intervention techniques, e.g.,, school staff members making personal contacts with businesses vs. school staff members working through the local Chamber of Commerce to achieve additional work experience stations for handicapped students. You will need to determine the statistical significance of the difference in effectiveness of the two techniques by applying a t-test analysis.

Step 6: Instrumentation

Every step is important. Each step must relate to each preceding and succeeding step. In selecting or developing instruments you must look back to Step 4 -- the point at which you decided what questions you need to have answered, and what indicators would provide those answers. Now that you are ready to select or develop instruments that will provide those indicators you must keep checking for congruence. Does the instrument really get at the precise information you need to answer your questions? Is it the simplest, fastest, least expensive way to get that information? Is it likely to produce accurate, valid and reliable information? These and other questions will help you make your instrumentation decisions.

For Stage 1, for instance, you need an accurate count of the numbers of businesses offering work experience for handicapped students at the end of the Stage 1 intervention. The instrument used to gather this information might be a structured telephone interview, a mailed questionnaire, a poll at a local Chamber of Commerce meeting. Which of these instruments would be mostly likely to yield accurate, complete, reliable information at least cost? None of these? You're probably right! So try again. Perhaps it would be best to develop a form for use by local businesses. The form could explain the purposes of the intervention effort and could provide space for the employer to record number, types, and duration of work experiences s/he was willing to commit. This instrument would be explained in the initial personal contact with the employer, and he or she would be encourage to complete it by a certain date.

Instruments for recording historical data related to work experiences for handicapped should provide for the data to be recorded in identical form, to avoid necessity for later conversion. For example, if the employer instrument asks for duration of work experience opportunities in terms of days, the historical instrument should also provide for recording days of availability.

Stage 2. To determine how many handicapped students participated in work experiences, for what length of time, in what kinds of activities, you will need another kind of instrument. You will also want to know whether this participation was due to your indirect intervention or to other variables such as changes in economic conditions, social changes, etc. Your instruments for gathering most of the data may be employer questionnaires and student questionnaires. Besides asking each to record participation, length of partici-

pation, and kinds of job activities, you might ask businesses why they made these openings available to handicapped students, and ask students why they participated in the work experiences. Responses to these latter questions will give evidence of the degree to which your indirect intervention was responsible for increase openings and increase participation. You will also need to analyze concurrent school and community programs to see whether any of them had a similar focus that may have influenced students' participation in work experience.

Stage 3. At this stage you need to determine how your efforts affected students. You have achieved more openings in the community, you have more handicapped students participating in work experiences in the community, but so what? Has it changed their attitudes toward work? Has it increased their

self confidence? Are their career aspirations more realistic? Have they developed career-related skills? Back in Step 2 when you were asking questions and specifying indicators, you should have recorded the types of attitude and skill changes you anticipated. After examining data from Stage 2 (types of experiences students received) you could refine your list of attitudes and skills you could reasonably expect students to develop as a result of these experiences. Now you need instruments to determine whether these attitudes and competencies have been achieved.

For attitudinal changes you may select a standardized instrument. Self concept scales are also available commercially. Or you may want to construct your own instruments to measure these aspects of student change. If you decide to construct your own, you should follow the suggestions presented in publications such as Anastasi. Model #5 in this Handbook gives some guidelines for evaluating development of products such as instruments.

To get at skill development you may need a performance test (e.g., if PBX competence was to be developed, a trained observer could observe a time sample of the student's PBX operations on the job); a form describing the demonstration task and the success criteria would need to be developed. Or you might need to use a rating scale to be completed by the employer. Rating scales for some skills are available commercially; local construction of such scales should be guided by reference to documents such as Model 1 in this publication and verified by someone expert in that skill. For each change in attitude or skill there must be an instrument from which it can be inferred that the attitude or skill has been acquired. A recent publication from Ohio State's Center for Vocational Education (McCaslin, Gross, Walker, 1977) describes currently available career education instruments.

Step 7: Data Collection and Management

If the credibility of your evaluation is to be established, you will need to design fail-safe procedures for data collection and management. During Step 4 you decided on the questions to be asked, and the indicators that would provide answers. In Step 6 you decided on the instruments to be used to collect the data. Now it is time to insure accuracy, timeliness, and completeness of data collected. You will need to determine: (a) the who and how of historical data collection; (b) the who and how of instrument administration; (c) the who and how of instrument scoring; and (d) the data reduction/summarizing techniques least likely to distort the data and most likely to provide information about differential results.

Again using our example of an indirect intervention to increase work experience options for handicapped students, you will approach data collection differently for each stage. For Stage 1 your principal indicators are numbers and kinds of work experience options available to students at the end of the project year as compared with historical data from the last three years. To insure accuracy of historical data, you will need to evaluate the various

sources of this information, and define procedures for searching those sources and recording findings. You will also need to choose for this task a person who has had experience in searching and reading historical records, who is known to be meticulous in data collection, and who is free from bias (has nothing to gain by focusing on some data while ignoring others). Timelines will be facilitated by establishing a reasonable timeline and displaying the timeline on the TASK/TALENT/TIME analysis chart (T/T/T -- see p. 87). Completeness can be assumed only if there are specific guidelines for the data collector. The data collector should be provided a form for recording the data, and for signing a statement that the guidelines have been followed and that the data are complete and accurate. Suggesting to the data collector that spot validation checks will be made may prevent sloppy procedures.

For Stage 2, indicators were numbers of handicapped students participating in work experience at the end of the second phase of the project, and types of experiences they had, as compared with numbers and types over the three years preceding installation of the activity. We have already suggested ways to insure accuracy, timeliness, and completeness of historical data. Here we will look at data collection for the project period.

You will need to design procedures and forms for collecting accurate, complete and timely data on: (a) numbers of handicapped students having work experiences in each of the two project years; (b) length of each of these experiences; and (c) types of experiences. Data may be collected from students and from employers; discrepancies in reports from the two sources will need to be resolved. Recognizing that both employers and students may be biased in their reporting, you will apply data reduction/summary techniques and arrange for data analysis; display and report only after you are sure your data are reliable and valid. If time, personnel, and funds are available and employers are cooperative, you may prefer to gather data by direct observation.

Space precludes our carrying the example through Stage 3. The suggestions for Stages 1 and 2 apply for any stage. However, there are a few more points to be considered in data collection and management. The first of these is consistency; if more than one person is collecting data using the same instrument, be sure each is following the procedures exactly. For instance, if one instrument administrator imposes a time limit and another does not, data from the two administrators will not be comparable and cannot

be combined. Do not depend solely on written directions for administration; hold a training session to review procedures and to impress test administrators with the importance of following those procedures.

Another caution centers around data reduction. If you are gethering data about unlike populations (e.g., one year the most common handicap may have been orthopedic, whereas another year it was visual), you will want collapsed data but you will also want to preserve data on students with specific handicaps; these latter data are most likely to provide support for decisions you will need to make. Again you are encouraged to look at the decisions you need to make and the questions you need answered in order to make these decisions -- then be sure your data reduction and analysis techniques do not lose the very information you need.

Step 8: Monitoring System

Great plans are rendered useless if you do not install and implement a meticulous monitoring system. The monitoring system provides for a breakdown of tasks required to implement each part of the plan, assignment of a person to be responsible for each task, and the time by which the task should be completed. This information is best recorded on a form known as a T/T/T -- Task/Talent/Time Analysis. The form decreases the likelihood that there will be slip-ups, slipped timelines, or confusion as to who's responsible for

what. A clerk can prepare a separate T/T/T for each project staff member, giving him/her an easily followed chart of the tasks for which s/he is

responsible. The personal T/T/T and the project T/T/T serve as roadmaps for implementing evaluation plans. Your T/T/T might look something like this:

For Stage 1: Increase number of businesses providing work experiences for handicapped students

	TASK	TALENT	TIME
1.	Design procedures for gathering historical data (numbers of businesses providing work experience for handicapped students during last three years).	Evaluator	by 9/15/79
2.	Recommend kinds of data to be collected from each local business to determine current opportunities	Evaluator	by 9/15/79
3.	Collect data for 1	Project Secretary	by 10/1/79
4.	Collect data for 2	Project Coordinator	by 10/1/79
14.	Prepare instruments for gathering post-treatment data	Project Coordinator	by 3/1/80
15.	Determine number of local businesses declaring willingness to offer work experiences to handicapped students	Project Coordinator	by 6/1/80
16.	Analyze data and prepare report	Evaluator	by 6/30/80

For Stage 2: Increase number of handicapped students participating in work experiences in local businesses.

	TASK	TALENT	TIME
1.	Design procedure for collecting historical data: number of handicapped students participating in work experiences during the last four years.	Evaluator	by 9/15/79
2.	Design protocols (written and oral interviews) for determining: (a) numbers of handicapped students participating in work experience during the current year; (b) length of each student's experience; and (c) type of experience.	Evaluator	by 11/1/80

3.	Gather data for 1.	Project Secretary	by 10/1/80
4.	Gather data for 2, etc.	Project Coordinator	by 6/1/80

For Stage 3:

	TASK	TALENT	TIME
1.	Design/select instruments for assessing attitudes and competencies developed by handicapped students participating in work experiences.	Evaluator	by 10/1/82
2.	Design procedures for identifying resources for developing these attitudes/competencies in the school setting.	Evaluator	by 10/1/82

These examples are not intended to define tasks for you; rather they illustrate the types of evaluation tasks that should be specified, and an effective method of specifying these tasks so implementation of the evaluation aspects of the program can be monitored. Although these T/T/Ts specify only evaluation tasks, you need to prepare a master T/T/T integrating all project tasks. As you do this you will quickly see the interdependence of tasks to be performed by different staff members. You will see which tasks must precede which other tasks, and what are reasonable chronological time lines for each task. Preparation of the master T/T/T should be accomplished cooperatively, with each responsible person having an opportunity to determine the appropriateness and the timeliness of the tasks for which s/he will be responsible. Initial design of the T/T/T should be a joint effort of the project coordinator and the evaluator.

A certain amount of rigidity appears to be implicit in the T/T/T. Some rigidity is intended; in other words, the inexorable passage of time is less likely to find the staff procrastinating if tasks, responsible persons, and time lines are recorded and monitored. However, even the most carefully thought-out plans may hit a snag when an unanticipated event -- so common in a human services organization -- intervenes. At that point, if a time line

becomes unrealistic you do not simply advance that time line; rather you survey all remaining tasks and time lines and determine whether several or all of them will need to be adjusted. The monitoring system is intended to help you manage your evaluation, not to manage it for you.

Step 9: Reporting

One of the most important functions of the evaluator is to display and report data in a manner that is objective, complete, and accurate. Such display and report should be honest and straightforward, and it should be easily understood by its intended audience. Data displays should be designed to highlight facts found, not to distort data in order to make a point. If results turn out not to fit one's expectation, then it is up to the evaluator and project director to generate constructive alternative activities that build upon the conclusions of the evaluation. Remember that the evaluation aims to provide information and that decisions based on information are nearly always better than decision based on ignorance or misinformation. Also remember that it is no distortion to point out ways in which contanimating factors may have affected the study's conclusions.

You will need more than one report. A technical report with every process step carefully delineated will be needed if there is a funding agency

involved. A decision-oriented report will be needed by the administration and the Board of Education; this report will highlight those findings on which recommendations are based. It is called an executive summary. An informational report will be needed by staff members and students; this will give evidence of the value of their efforts. And finally, a public report for community and press will describe what was done and what effect it had. Language of each must be appropriate for the intended audience if communication is to result. None of these reports needs to be lengthy. Each should speak directly to its audience. Following is an example of an executive summary related to our model.

PROJECT "ACCESS"

Executive Summary

State legislation requiring equal educational opportunities for handicapped students, and recent trends toward mainstreaming students formerly assigned to special classes, stimulated a reexamination of the educational goals for handicapped students at Porpoise High School. An analysis of the relationships between teachers' expectations of handicapped students and of nonhandicapped students of comparable academic ability, revealed substantially lower expectations for the handicapped. Analysis of achievement grades showed that handicapped students' grades reflected lower performance levels than did the same grades given to nonhandicapped students. Handicapped students tended to be afforded fewer diverse experiences than nonhandicapped students, with handicapped being confined largely to vicarious experiences (e.g., reading about, rather than doing).

In order to determine which differences in expectations and experiences for handicapped and nonhandicapped students were most in need of adjustment, an instrument was developed, pilot tested, field tested, and finally administered to all handicapped students and to a random sample of nonhandicapped students, parents of handicapped and nonhandicapped students, teachers, counselors, administrators, and members of the business/industry community. The instrument was accompanied by an explanation of the handicapping conditions, emphasizing areas of strength, and of the need for handicapped to be absorbed into society upon exit from school. Findings of this survey showed overwhelmingly that whereas handicapped students have some special needs, they also share nearly all the needs of their nonhandicapped peers.

To determine extent of discrepancy between work-related attitudes and skills of handicapped vs. nonhandicapped students, a list of basic work competencies was prepared, and an instrument for measuring these competencies was developed, using items from the Career and Occupational Development Tests of the National Assessment of Educational Programs. All handicapped students and a sample of nonhandicapped students matched by age, sex, grade and ability were administered the test. Results were overwhelmingly in favor of the nonhandicapped students. Analysis of the areas in which the greatest discrepancies appeared showed that nonhandicapped students were lowest in effective work habits, attitudes toward work, and preparation for making career decisions. After examining alternative programmatic efforts, it was determined that these

attitudes and skills could best be developed through community work experiences.

Analysis of the work experience records of all Porpoise High School students revealed that very few handicapped students had ever participated in school-related work experiences in the community. The staff designed PROJECT ACCESS to create work experience stations in the community for handicapped students.

Method

It was decided that instead of working directly with the handicapped students, trying to teach them the skills they lacked, it would be more propitious to employ an indirect intervention working with community businesses and industries to make them more receptive to handicapped students on the job, and hopefully to encourage them to open work experience stations to handicapped students. Work experience records of the last three years were examined, and work stations for these years were tabulated by type of work by time on the job, and by student characteristics. This furnished baseline information against which work experience stations and experiences available after the program could be compared. Forms were developed for project staff members to record specific information regarding contacts with community employers. A protocol was developed for these contacts. The evaluator instructed interviewers in the use of the protocol and forms. A T/T/T was developed to monitor project tasks.

At the end of the project year, project staff conducted interviews with employers who had provided work experience stations for handicapped students, and with handicapped students who had been assigned to these stations. Interview protocols and forms provided for gathering of information about the number and nature of work experience stations available, the number and characteristics of the students assigned to these stations, length of assignments, and the factors that influenced students to seek these assignments.

This information was compared with baseline information, and it was found that the number of work stations for handicapped students had increased from 17 to 42 and that the number of handicapped students participating at these work stations increased from 119 to 275. As shown on the next page, this indicates a significant increase in percentage of handicapped students participating. A *t*-test of the increase in percentage (two tailed) was significant beyond the .0001 level. The number of participants more than doubled, so that instead of fewer than half the handicapped students participating, now more than five out of six were. The average length of assignment to work stations

	Prior Year	Target Year
Number (%) of Students Participating	119 (39%)	275 (85%)
Number (%) of Students Not Participating	183 (61%)	48 (15%)
Total Handicapped Students		

$$t = \frac{(P_2 - P_1)}{\sqrt{\frac{P_2(1-P_1)}{N_2} + \frac{P_1(1-P_2)}{N_1}}} = \frac{(.85 - .39)}{\sqrt{\frac{.85x.15}{323} + \frac{.38x.61}{302}}} = 13.4$$

for the 323 handicapped students was 4.1 weeks (s.d. = 1.2 weeks), and for 312 nonhandicapped students it was 4.4 weeks (s.d. = 1.4 weeks.) Although this difference was statistically significant at the .001 level,* it represents only a 1-1/2 day difference out of over 20 days, or less than an 8% effect on the average. We do not believe it to be an important difference.

We examined alternative possibilities for causes of the dramatic increase -- such as a different mix of handicaps from one year to the next, external influences on businesses to participate and provide work stations, or a false classification of some experiences as real work assignments when they were really only perfunctory or nominal. We could find no evidence to diminish our belief that Project ACCESS efforts had actually caused a dramatic increase in work exposure for handicapped students.

Examination of the handicapped students' identification of factors that influenced them to seek the new assignments revealed that no direct intervention had been needed; the fact that the stations had been created and had been made known to the students appeared to be enough to facilitate their access to the work experience stations.

Comparison of the characteristics of handicapped students who had not taken advantage of the work experience stations with those who had, revealed no differences in handicaps or in general abilities, but significantly large differences in attitudes toward work and in work habits.

---

* $$t = \frac{4.4 - 4.1}{\sqrt{\frac{322(1.2) + 311(1.4)}{633}\left(\frac{1}{323} + \frac{1}{312}\right)}} = 3.3$$

Follow-up Implications

Since the indirect intervention was successful in achieving a significant increase in numbers of community work stations available to handicapped students and in numbers of these students assigned to such stations, it appears that this is a feasible and effective program, and that it should be continued.

The differences between handicapped students who had and who had not availed themselves of the newly created stations indicated that direct interventions may be needed to help some students gain those attitudes and skills needed to access the work stations. Project staff plan to mount such a program during the next school year.

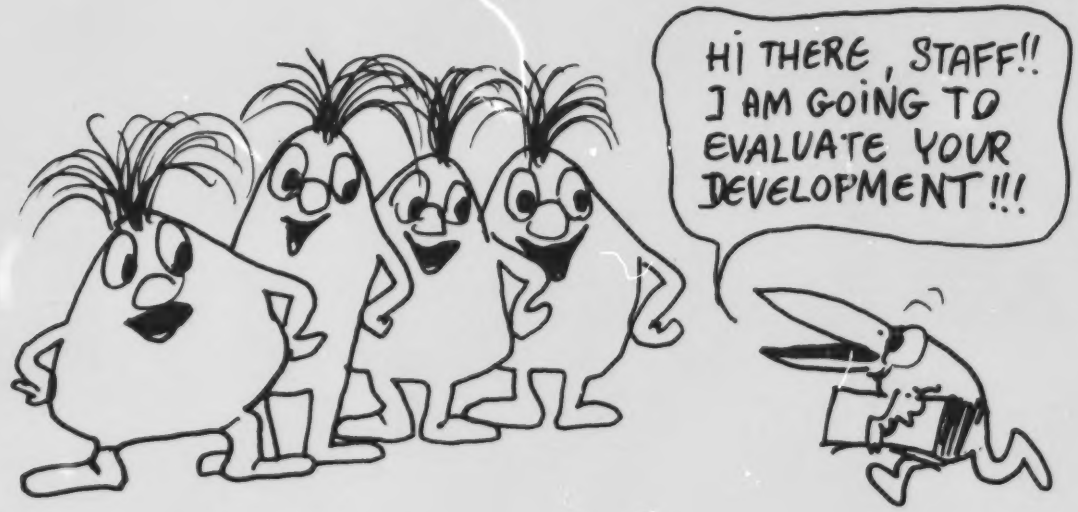
Project staff also plan to share with employers a list and explanation of the specific work-related attitudes and competencies handicapped students are expected to achieve while participating in the community work experience, and to study the costs of such competency development during the next school year. Such cost and effectiveness data will help the staff make decisions about the desirability of continuing to expect work competencies to be developed on community work experiences, or whether it would be more parsimonious to provide for such instruction in the school.

EVALUATION OF STAFF DEVELOPMENT

Model #4

Octave V. Baker
Jack A. Hamilton
Donald H. McLaughlin

Inservice staff development activities play a central role in the development of career education programs because career education is usually viewed as an integral part of the educational process. Teachers with well-developed styles, procedures, and curriculum plans have a right to ask, however, whether career education is worth the investment of their time in staff development. Evaluations of staff development activities should aim to help these teachers select appropriate development activities. That is not the only purpose for the evaluation, however; career education coordinators and others planning staff development activities need to know whether their activities are achieving the goals and filling the needs for which they were designed. Otherwise, there is no impetus to improve those activities. And, of course, there is the ever-present necessity for showing the value of the efforts by securing continued support.



How you set out to conduct your evaluation depends on the type of staff development activity you are evaluating. As shown in Table 1, different career education staff development activities cover different topics and use different methods. (The entries in Table 1 show the relative frequencies of each combination found in a recent national survey.) Of course, many activities cross over and cover several topics, but the objectives against which performance is to be evaluated can be classified according to the topics--and the evaluation may show, for example, that an activity successfully conveys concepts but only weakly improves teaching techniques. Each topic has its particular mode of evaluation. Concepts represent usable knowledge and can best be tested by administering problems that require their understanding for solution. Teaching techniques can best be tested by observing participants' performance in teaching contexts. Curriculum materials objectives can best be tested by measuring effectiveness of materials developed, selected, and/or adapted. Community resources objectives can best be tested by measuring performance in situations where

Table 1

Percent of Students in Districts where Various Types of Career Education Staff Development Activities Occurred

	Career Education Topics				
Methods	Concepts	Teaching techniques	Curriculum materials	Evaluation techniques	Community resources
Internal staff activities					
Presentations by local school staff experts	40	35	32	20	26
Group sharing sessions among school staff	37	31	30	19	23
Locally prepared staff development materials	34	33	38	23	31
External educational resources					
Attendance at professional meetings	42	28	28	17	15
Published staff development materials	36	28	35	17	16
Presentations by State Career Education experts	36	27	28	20	16
Presentations by experts from institutions of higher learning	31	24	21	15	15
Visits to other Career Education sites	29	23	23	13	14
Summer sessions at college	23	18	17	10	10
Workshops by professional organizations	22	16	16	9	9
External non-school resources					
Visits to business/labor/industry or other community sites	31	16	18	7	30
Presentations by experts from business/labor/industry	23	14	18	5	25
Work experience in fields outside education	16	9	10	4	11
Workshops by private firms	8	8	8	4	6
All methods combined	66	67	67	66	65

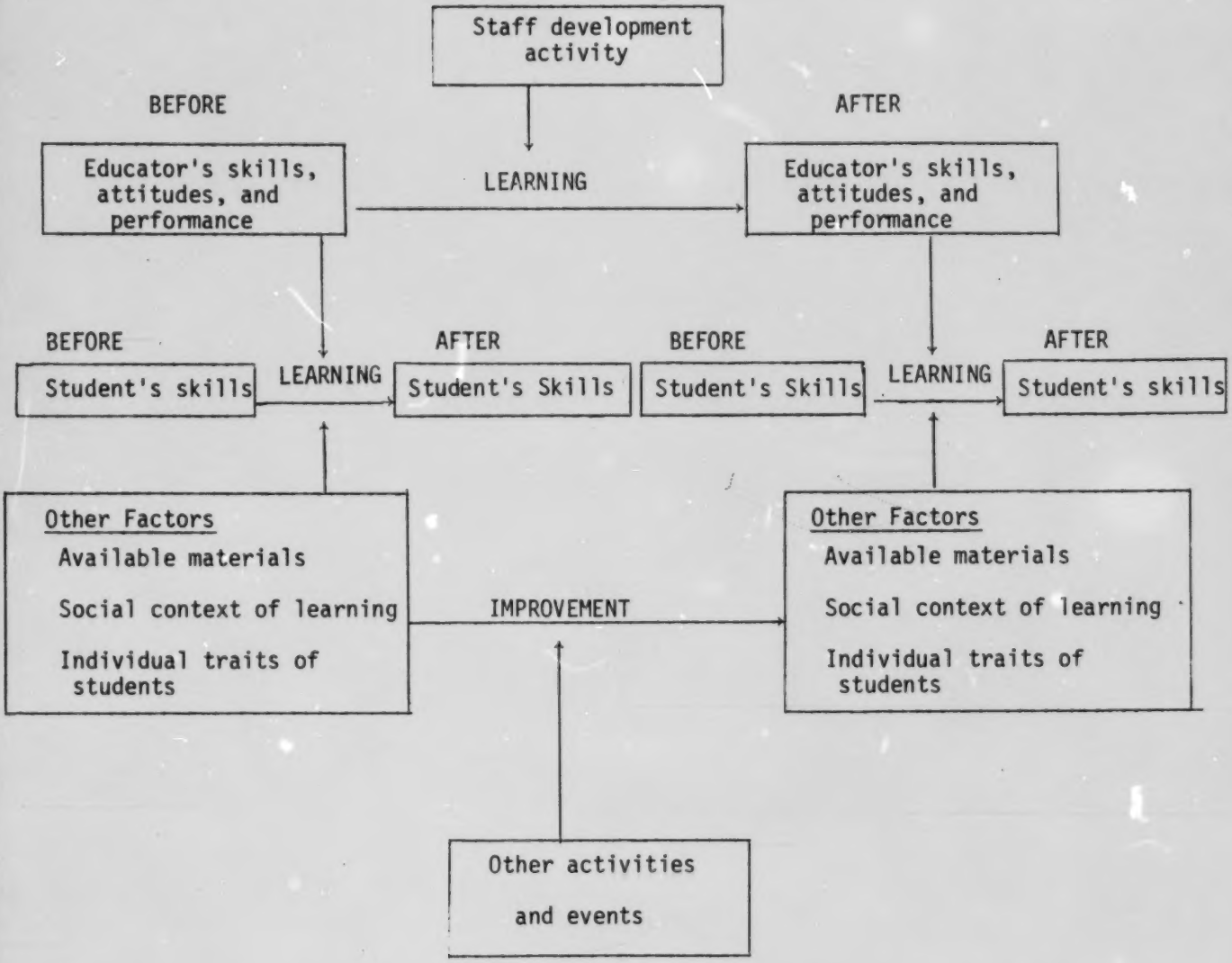
community resource utilization is appropriate. Evaluation objectives can best be tested by observing the conduct of an evaluation by the participant. Of course, testing for the attainment of these objectives by actual performance observation is itself costly, and simpler operations must generally

substitute. One useful substitute is to measure the participants' ability to critically appraise the performance of a well-chosen model (e.g., using questions like "What's wrong with this teaching technique?").

The model we shall describe is based on the premise that what is being evaluated is merely the staff development activity, not the total career education program. Quite cogent arguments can be made that students' needs for career development skills are an ultimate reason for staff development activities in career education and that their needs are to be taken into account in deciding on those activities. As shown in Figure 1, however, that is a causal leap that cannot be easily made because many other factors besides staff development activities affect students' career development. Unless you are willing to assume that no other activities and events affected student learning rates over the same interval or unless you know what those effects are, you cannot confidently attribute gains in student learning (or lack thereof!) to the staff development activity. If you measure gains in the skills of the participants in the staff development activity, on the other hand, you are in a position both to evaluate the implementation of the career education staff development activity and also, by using correlations, to relate staff competencies to student learning. If, in evaluating the overall effectiveness of your career education program in terms of student gains, you find greater benefits for the students of those teachers who gained most from the staff development activity, you will be able to document the importance of staff development in your career education program. In order to help you do this, we have included a discussion of correlations at the end of this chapter.

Staff development activities always require some investment (money, efforts, materials, participants' time, etc.) and one purpose of evaluation can always be seen as to determine whether these investments paid off. More should be expected from expensive activities than from less costly ones. As can be seen in Table 1, there are numerous methods for staff development activities, each involving particular inputs, and when a choice of methods is to be made, the results of evaluation of a previous year's cost and benefits can help to make a better choice now. It helps in designing your evaluation and choosing the questions to ask to (1) remember that the results may be used in next year's choice, and (2) consider what alternatives you will be choosing among them. Some of your participants will have had prior experiences in staff development activities using different methods, and if you ask them the right questions, they can help you to compare different methods.

Figure 1. General Diagram Relating Staff Development to Student Skill Achievement



In making cost and benefit comparisons, which are useful even if they are inexact, you will need to consider many different kinds of costs--costs to the school budget, costs borne by the participants (including volunteered time), and costs borne by the students through loss of interaction with teachers (if the activities are on school time). Consideration of these costs is likely to affect the decision of when to hold staff development activities. Thus data you gather on costs now may play a useful role in next year's decision.

Decision - Rationale - Information Need Approach to Evaluation

What we have been saying is that your evaluation is useful just to the extent that it helps in making decisions. Rather than ask a standard set of questions and include average responses in your report about a workshop, it will be better for you to consider ahead of time the decisions you want to facilitate. Then you can focus on ways to get participants to provide information to help you evaluate arguments to be made in the future favoring one choice over another.

There are five steps in planning for an evaluation this way: (1) identification of decision makers, (2) identification of decisions, (3) identification of arguments (rationales) favoring different choices, (4) identification of information needed to test those arguments, and (5) figuring out how to get the information. Most evaluation models focus on

Step 5, which is the most complex, so we think it is a valuable counterpoint to discuss the first four steps here. In particular, what we present here does not differ dramatically from an application of Model #6 (Evaluation of Curriculum Implementation) to staff development activities, but here we focus on the first 9 of the 18 tasks presented on page 165 of this handbook.

Step 1: Identify Decision Makers

These are the people you hope will say, after seeing your evaluation results, "This helps me decide which way to go!" To determine who they are, first list all the categories of individuals who might be expected to participate in some way in career education. You should do this even if you know you are doing the evaluation for a particular audience, such as a state funding program that requires evidence of evaluation. That is because the more people you can serve with a single data collection effort, the more valuable your evaluation activities will be seen.

Different categories of decision makers to consider include:

- career education coordinators
- school superintendents
- school boards
- prospective participants (teachers, counselors, other school staff)
- presenters of staff development information
- parent advisory groups
- administrators in other school districts
- outside funding agencies

Of course, in your particular situation some of these may play only limited roles, and there may be other decision makers to consider, such as particular deputies of the school superintendent. It is important to get to know these decision makers, through interviews or at least by telephone contact, prior to your evaluation. The content of your conversations will help complete Steps 2, 3, and 4. Note, however, that you must begin these steps before contacting the decision makers.

Step 2: Identify Decisions

One way to identify decisions is to talk to the decision makers. This is not fruitful, however, if you are not prepared. People are sometimes hesitant to describe their decision-making roles and sometimes even unaware

of or unclear about their roles. Quite often the distinction between "decisions" and "recommendations for decisions" is vague and may depend on such things as the wording of the recommendations or how busy the person is who reads the recommendations. It is therefore best for you to have in mind a set of decisions the person might be involved in before talking to her/him. Then you can add to this set any new ones that come out of the conversations.

There are basically three types of decisions to be made concerning a career education staff development activity, although they become different decisions for specific decision makers:

(1) whether to provide support for the activity,
(2) whether to participate in the activity, and
(3) how to change the activity.

As the third of these types of decisions implies, we are assuming that the decisions are made after the activity has occurred at least once. You might say that evaluation cannot help in an initial decision and you would be right, but few decisions are "initial." In developing any new activity, you will do better to become aware of existing activities elsewhere and to use whatever evaluation data are available about them to help guide your development. Looking at evaluation in this perspective you can see that evaluating your ongoing program well can help countless others in the future who come across your results. We focus our consideration here, though, on the case where it has already been decided that a particular career education staff development activity is to be implemented; your evaluation is aimed at decisions for the future of that activity.

The decision <u>whether to provide support for the activity</u> may be made by a career education coordinator, who might decide to allocate efforts elsewhere, or it might be made by a school superintendent or his/her deputy in deciding on a career education program budget, or by the school board faced with budget allocations. It may be made outside the school district, by a funding agency seeking to promote creative and successful approaches to career education. It may also be made by teachers, in the form of recommendations on school priorities and by parental and community advisory groups. In each case, you need to know what <u>alternatives to supporting the activity</u> they are considering. These are the alternatives against which

the activity will in reality be measured, and to be convincing you must design an evaluation that might come out supporting the decision to abandon the activity. If this requires your hiring an outside evaluator, do it - if you cannot honestly agree that you would be convinced by the data if they show your activity does not merit support.[1]

The decision whether to participate in the activity primarily applies to school staff and implies that participation is voluntary. It is especially important to consider school staff in choosing your evaluation objectives because they will provide most of the evaluation data. In order to be fair, any evaluation should provide value to those who contribute to it. Moreover, if the participants are to be actively involved in the conduct of the evaluation, as we discuss later, they must be able to understand the relations of questions to decisions to be made, and the decision for them whether to participate involves rationales and information needs very similar to those involved in decisions by administrators whether to support the activity.

Decisions on how to change the activity are easier to make when there are real alternatives to consider. Merely deciding that a change is necessary because objectives are not being met does not lead efficiently to deciding ona particular alternative. Thus, if you find that objectives were not met, you will provide a much greater service if you simultaneously report on specific processes that were related to the failure. This requires, however, that you develop a model of your career education staff development activity that shows the relations of various processes to outcomes and the effects of various positive and negative factors on those relations. Suppose you are evaluating group discussion sessions. Positive factors might include specific relevant prior experiences of some of the participants; negative factors might include participant absences from critical sessions. These factors could substantially alter the effectiveness of group discussions, and obtaining information about them in your evaluation will greatly increase the value of any conclusions you reach concerning the value of group discussions.

---

[1]Note we are writing this assuming you are actually a career education supporter, not an unbiased observer (if such exists). In evaluation, any evaluator must set aside her/his personal biases (and therefore must be aware of them) and plan to collect data that would be viewed as valuable from various perspectives, even those she/he disagrees with, if players in the decision-making arena hold those perspectives.

Step 3: Identify Rationales for Decisions

This is the critical step in relating data to decisions. You must develop an idea of the arguments or rationales that will be used by decision makers to arrive at and support their decisions. In doing so you must be aware that there are some arguments for which evaluation is irrelevant, such as the continuation of a program without change as a reciprocal act for somebody else's agreement to support another program. These compromise arrangements may actually lead to smoother and more efficient school system operation. Having identified these possibilities, however, you need not abandon evaluation, because the data once collected can have indirect impacts. For example, evaluation can make it more difficult to defend poor programs that continue to operate and thus can ultimately cause them to be changed.

For the decision whether to provide support for an activity, we can identify several rationales pro and con:

Rationales for support:

(a) There continues to be a need which the activity satisfies or aims to satisfy.

(b) The activity is meeting its stated objectives.

(c) The activity is more cost effective than other alternative(s) under consideration.

(d) Discontinuing the activity would have negative side effects.

Rationales against support:

(a') A need which the activity satisfies or aims to satisfy no longer exists.

(b') The activity is not meeting its stated objectives.

(c') There is another activity which is more cost effective.

(d') Discontinuing the activity would have positive side effects.

No single rationale should be expected to control a decision if there are competing arguments on the other side. Your aim must be merely to collect valid evidence about the accuracy of the arguments. For example, without valid evidence it is possible for proponents to argue from individual cases that an activity is meeting its stated objectives and for opponents simultaneously to argue the reverse. Only an evaluator's data can eliminate such weakly based arguments by showing the overall extent to which the activity is meeting its objectives.

For the decision whether to participate in an activity, there are also several arguments pro and con:

Rationales for participation:

(e) I am having problems integrating the concepts from the world of work into my role as an educator.

(f) Educators who have participated are able to do things I want to do.

(g) This is an easier way for me to gain the skills than another activity I was considering.

(h) There are external reasons I should participate.

Rationales against participation:

(e') I am able to integrate concepts from the world of work into my classroom as well as I want to.

(f') Educators who have participated seem to be no better for the experience.

(g') There is an easier way for me to gain the skills.

(h') There are external reasons I should not participate.

Each of the rationales for individual participation can be seen to be a personalized version of the corresponding administrative rationale for providing support for the activity.

For the decision how to change an activity, the rationales are somewhat different:

Rationales for a particular change:

(i) A particular process did not occur as planned, and apparently as a result the outcome was less beneficial.

(j) A particular alternative to a process involved in the activity would produce a more beneficial outcome.

(k) A particular process had an unfortunate side effect.

(l) Changing the process would lead to other changes that would improve the outcomes.

Rationales against a particular change:

(m) Change per se is costly.

(n) The outcome now is at least adequate, and the outcomes from alternatives are unknown.

(o) The side effects of a process justify its cost.

Improving a program through incremental changes is one of the most valuable outcomes of evaluation. It also requires more creativity and research on the part of the evaluator: having found a problem s/he must generate alternatives and search for data that would suggest whether or not the alternatives would be any better.

The identification of decisions and rationales should be undertaken in iterations. First, using common sense, generate a list of decisions and rationales, like those listed above. Next, present these informally to some or all the real decision makers you identified in each category and ask for their serious critique. From this you should obtain some new ideas while at the same time creating images in the minds of decision makers about the evaluation's outcome and uses. This may be somewhat threatening to some decision makers who may see it as an infringement on their decision-making power. It is possible to reduce the threat involved in evaluation by pointing out positive aspects of each decision alternative. Finding a staff development program in trouble, for example, suggests an opportunity for exploring new methods. By properly preparing for and conducting interviews with decision makers, an evaluator can greatly increase the value of an evaluation to the school district and to the career education program.

Step 4: Identify Information Needed to Test Arguments

It is straight-forward to transform decision-rationales into information needs. You merely ask, "What evidence do you have to support your argument?" Evaluation aims to provide the information to answer that question or to

demonstrate that the evidence does not support the argument. In this section you will find a very general description of the information needed for testing each of the rationales listed in the preceding section. You will find it useful to refine these information needs by asking decision-makers just what information would convince them. You may actually find their standards much less rigorous than your own, and you will have to make your own judgment about rigor. On the other hand, you may find some individuals difficult to satisfy, and adding other measures to your evaluation plan in order to satisfy them may make a substantial difference in their use of the evaluation results.

Rationale (a) (for/against support): There continues to be (or does not continue to be) a need which the activity satisfies or aims to satisfy.

Information Need: The information needed for this rationale-pair does not involve the staff development activity directly but is an assessment of the skills, attitudes, and performance of educators who might be prospective participants. If your activity's objectives include specific performance criteria, then this assessment could be carried out using the same measuring instruments one would use in assessing whether the activity is achieving its objectives.

For example, if your career education staff development activity aims to inform teachers about the stages of career development, an item requiring listing and describing developmental stages would serve both to assess attainment of the objective and as an indicator (for nonparticipants) of the need for the activity.

Rationale (b) (for/against support): The activity is meeting (or not meeting its stated objectives.

Information Need: In order to evaluate this argument, it is necessary to translate objectives, however broadly stated, into observable events. In the translation process a part of the generality of the objectives is bound to be lost, and only by carefully considering the whole domain of skills, attitudes, and/or performances the activity is designed to affect can you select a sample for measurement that will generalize to the whole domain. For example, if you used one occupation as an example in presentation but want to assess the acquisition of the general concept, your test item should refer to a different occupation.

The expected behavior can involve cognitive, affective, or psychomotor outcomes.

- Cognitive objectives refer to knowledge outcomes.
- Affective objectives refer to outcomes involving expressed attitudes and feelings.
- Psychomotor objectives refer to skill outcomes involving some type of physical action.

The key question is: Did your program of staff development activities produce the kind and amount of results you anticipated?

To illustrate this step, an example can be drawn from the Maryland Career Guidance Institutes, an extensive staff inservice training program whose goal was to provide participants with the skills and competencies enabling them to implement career education and strengthen career guidance in their schools. In preparation for the institutes, Maryland State Department of Education personnel constructed a detailed set of desired post-training outcomes. The outcomes described 19 skills and competencies participants were to possess after their Institute experiences. The 19 skills and competencies were grouped into eight categories, as follows:

- Conduct group career guidance activities
- Systematically facilitate student career planning and decision making
- Assist students in assessing career-related attributes such as abilities, interests, and attitudes
- Plan a placement program (secondary school level)
- Plan a school career exploration program (middle or junior high school level)
- Plan a prevocational educational program (secondary school level)
- Plan and organize a school career resource center serving the community as well as the school
- Plan for the implementation of career guidance elements for a comprehensive career education program

In addition to these knowledge outcomes, the planners also specified that participants should come out of the Institute with improved attitudes and motivations toward activities in the eight categories. Thus, affective outcomes were also considered in the development of measures for the evaluation.

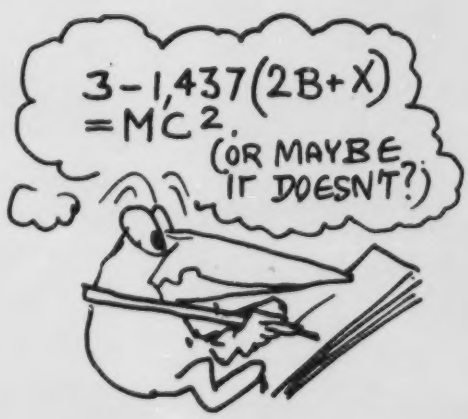
In order to develop the clearest picture of how well each objective is being achieved, you should consider several different measures related to it. You may wish to state an objective as both (1) a postactivity skill level and (2) a pre-postactivity gain in skills. A significant gain more clearly shows the impact of the activity, but reaching a particular postactivity level more clearly shows that the participants are prepared for the career education roles the activity was designed to teach.

Another important variant is to validate the primary post-activity measure of attainment by later measurement of performance in the school setting. This may be done on a random sample of the participants to limit costs. A description of the use of correlation coefficients to validate measures is appended at the end of this chapter.

Rationale (c) (for/against support): The activity is more (or less) cost effective than some other alternative under consideration.

Information Need: If you are lucky enough to have a program of sufficient size to try out two different activities or if your activity is comparable (has the same objectives) to one for which cost and effectiveness data are available, you may be able to make cost-effectiveness comparisons. If the less expensive program turns out to be the more effective, your results will be easy to translate into a decision, but if the more expensive program is more effective, it is more difficult. You either need (1) to estimate how much more it would cost to bring the less expensive program up to the effectiveness level of the more expensive one, or (2) to estimate the gain in effectiveness that would result from equating the costs by increasing the scope of the less expensive program. If neither of these is possible, then the rationale cannot be adequately tested.

In calculating costs you need to keep in mind the decision to be served--the cost of implementing the activity one time may be quite different from the next time. Cost-effectiveness applies to the activity you choose to implement again, not to the activity you just completed. In calculating costs you also need to include factors that may not have



easily obtainable dollar values, such as school space or volunteered time. If some needed resource will be free to you, you need not include it in your cost computations for decision-making, but you will mislead others reading about the activity if you don't mention it as a needed resource in your report.

Rationale (d) (for/against support): Discontinuing the activity would have negative (positive) side effects.

Information Need: Side effects, by definition, are related directly to the stated objectives of the activity, and it is not possible to devise a questionnaire to cover all possible side effects. Therefore, we recommend an open-ended approach in which you ask all participants to tell you in their own words what factors actually led to their decision to participate, what outcomes of the staff development activity, other than those clearly intended, they perceived, and what reasons they think the activity should (or should not) be continued. Addressing this last question to the decision-makers interviewed at the beginning of the study should also produce suggestions of side effects. Once potential side effects are identified, to the extent that they are either valuable or costly, they should be included in the data collection process.

A systematic approach to the identification of side effects is first to list categories of individuals, as in Step 1, and then to go over various types of behavior of each that might be changed because of the staff development activity. Common side effects include:

- the broadening of participants' communication patterns,
- the formation of attitudes about components of the method used in the activity,
- the effects of having to make the decision whether to support (or to participate in) the activity, and
- increasing the visibility of career education.

Each of these can be either positive or negative--or both--depending on the particular context.

Rationales (e), (f), (g), and (h) (for decisions whether to participate) are roughly parallel to the first four, and we need not discuss them at length here. They do offer a chance to place participants in the decision-making role so that they can better appreciate the purposes of evaluation. If time permits, you can convey evaluation concepts to the participants by asking them to go through the steps of: (1) identifying, possibly by interviewing each

other, arguments for and against participation, (2) identifying the information needed to test those arguments, (3) figuring out how to gather that information, and (4) figuring out how to convey that information to other prospective participants in the future. This will increase the participants' commitment to providing information, and that information is essentially the same as needed to fill other evaluation information needs--to assess existence of a need for the activity attainment of objectives, cost effectiveness, and side effects. Of course, if the staff development activity is a one-day workshop, very little time can be devoted to digressions from the main theme. Even then, however, spending a half hour focusing on the participants' reasons for participating and on the assessment of the value of participation may also reap benefits in increasing the participants' orientation toward the activity's goals.

Rationale (i) (for change): A particular process did not occur as planned, and apparently as a result the outcome was less beneficial.

Information Needed: In order to test this rationale, it is necessary to have a conception of the relations between the activity's processes and its outcomes; that is, a conceptual model of the activity. Without such a model, it is purely guesswork to attribute variation in outcomes to deviations from planned processes. Quite often actual processes will deviate from plans in order to accommodate and to take advantage of peculiar aspects of the situations surrounding the activity, and automatic conclusions that such deviations contribute to failure to reach specified objectives will only lead to reducing the flexibility of the activity. When deviations from planned processes occur, it can be crucially important for the evaluator to assess, possibly through impromptu interviews, the immediate outcomes of the processes. For example, the late arrival of materials for a workshop session may be unimportant, but if time ran out before the topics of the session could be covered because of the materials delay, the importance of that delay would be increased.

In order to fill this information need it will be necessary not only to administer an assessment instrument to participants but also to obtain a detailed record of the entire conduct of the staff development activity by an impartial observer. The record can then be used in interpreting the other results of the evaluation.

Rationale (j) (for change): A particular alternative to a process would produce a more beneficial outcome.

Information Need: Information is needed on the alternative process as well as on the process as implemented. Unless you can set up an activity with alternative processes to compare, you must rely on past experience and the experience of others to provide the needed information. Many of your participants will have had such experiences and they should not be overlooked as an information source about alternative processes. Ask them, "Were you ever in a session where they...? If so, do you think it was better or worse than our approach of..., and why?"

Rationale (k) (for change): A particular process had an unfortunate side effect.

Information Need: You need essentially the same information as for Rationale (d), the only difference being that here you are concerned with a particular process, not the entire staff development activity.

Rationale (l) (for change): Changing the process would lead to other changes that would improve the outcomes.

Information Need: Here again, as for Rationale (i), you need a conceptual model of the entire activity, and with that you need evidence that the "other changes" would be likely to improve the outcomes.

Rationale (m) (against change): Change *per se* is costly.

Information Need: You need to develop a plan for implementing the change and itemize the costs of the steps provided for in the plan. Quite often this process will lead to the identification of equally effective but less costly changes.

Rationale (n) (against change) The outcome now is at least adequate, and the outcomes of the alternative are unknown.

Information Need: As for Rational (j), you need information on the outcomes of the alternative. Do not overlook participants as sources for information on alternatives. This requires that you be aware of alternatives that may be suggested before you gather the evaluation data. Interviews with decision-makers will help you prepare for this.

Rationale (o) (against change): The side effects of a process justify its costs.

Information Need: This, like Rationale (k), is a variant on Rationale (d), focusing on a particular process rather than the entire activity. Therefore, the information needs are those discussed under Rationale (d).

When you have considered each of these rationales for decisions and made initial tentative plans for assessing them, you are ready to discuss the evaluation with decision-makers in order to refine your plans. As an outcome of the discussions, you should generate a memorandum to these individuals stating what you understand to be the most important decisions and decision rationales and briefly how you plan to address them. Attending to feedback from this memorandum will smooth your implementation of the evaluation.

Step 5: Figure Out How to Get the Information

Having discussed steps 1 through 4 at length, we will only briefly cover this step, which is covered in many places including other models in this handbook. We shall focus here on details that may be peculiar to the evaluation of staff development activities.

There are five sub-steps in the planning of information gathering: (1) gaining approval to use various information sources, (2) developing a schedule and allocating staff, (3) generating forms and data collection procedures, (4) planning for the collection of data before, during, and after the staff development activity, and (5) preparing procedures to use the data collected to evaluate decision rationales.

Gaining approval for using information sources. First, it is necessary to coordinate the evaluation activities with the main staff development activities so that evaluation helps and does not hinder progress. This can best be done by focusing on ways that evaluation and feedback can help the participants as they progress and by avoiding interruptive evaluation activities during critical periods in the staff development activity.

Second, it may be necessary to gain permission to survey other teachers to produce a needs assessment. Third, acquisition of comparison data from other sites may require several days, weeks, or even months of contacts and delays. Fourth, access to cost data for cost effectiveness analyses may require approval.

Developing a schedule and allocating staff. One way to ensure that you will run into difficulties is to plan only a week ahead. You should have a set plan that will get you from here to there. You should be flexible by considering two explicit ways of performing some subtask, not by saying, "We'll decide what to do when we get there." For each subtask in the evaluation, you need to run through it in your mind, figuring out how much time it will take (multiply by 2), what barriers may arise and how they can be dealt with, and what expertise is required to complete the subtask successfully. Asking an experienced evaluator to critique your schedule would not be a bad idea.

For example, if you are collecting responses from 100 participants, and if you included several open-ended questions requiring careful reading and categorization, then several full-time professional days will be needed to make best use of the answers given. Allowing less time will not be fair to the participants who took time to respond thoughtfully.

Generating forms and data collection procedures. Your data collection will primarily gather information on the skills, attitudes, and performance of participants. In every case, you should start by listing the behaviors and events in the school which the staff development activity is designed to increase and decrease. You would then count up the frequency of these events before and after the staff development activity, but the cost of such observations is likely to be prohibitive, both in terms of effort and in terms of delay. The target events may occur so rarely that trends over many months must be observed to reach stable conclusions.

Ideally, evaluation of the effectiveness of a staff development activity would pursue the extent to which the participants use the newly acquired knowledge/skills/behavior in the classroom, and the effect this has on students. In some instances, (particularly in the knowledge realm), e.g., teaching teachers how to relate each subject to occupations, measures of the activity's effects on teachers might include relevant bulletin board displays, classroom observation, lesson plans; measures of effects on students might include samples of students' responses to relevent assignments, and class discussion.

The next best thing to actual observation is simulation. Give participants situations where target behaviors and events should occur (e.g., the use of some teaching techniques) and observe and score their performance in these situations. Role-play simulations may be too time consuming; simulation

games which require participants to make a series of decisions consistent with the focus of the training may be pursued by a number of participants simultaneously.

A third alternative is to use paper-and-pencil items asking for responses requiring the knowledge, skills, or attitudes that the staff development activity was designed to produce. It is unlikely that you will find instruments appropriate for evaluation of your staff development activity, as such activities usually are specific to the needs of the individual school program, staff and students. Therefore you probably will need to develop your own instruments. However, the instruments need not be long or complicated. Just be sure that the item(s) designed to test each desired knowledge/skill/attitude present demonstration tasks from which you can infer that the behavior has been learned. For instance, if you ask participants to name three of the Office of Education occupational clusters, you cannot infer from their responses that they can name all 15.

Since staff development activities frequently are single sessions, or closely spaced consecutive sessions, pre and posttesting of the participants usually will give a fairly accurate picture of the degree to which the target knowledge/skills/attitudes have been transmitted. In the single session, the participants are insulated from outside influences for the duration of the training, and any changes in their performance can be ascribed to the staff development activity. Their use of the knowledge/skills/attitudes will require monitoring over time.

In addition to direct measures of objectives-attainment, you should remember the purposes of your evaluation, to test rationales for decisions, and you should include questions that directly assess the participants' ratings of those rationales. "Why did you choose to participate?" "What type of participant would benefit most from the activity?" "Which objective was least well met?" "What are some other good methods we might have used?" "How should this be changed next year?" Professional evaluators will criticize an evaluation based solely on such questions, but as a supplement to "hard" evidence about gains in skills, answers to these questions will help program planners and other teachers make accurate choices in the future. The criticism of such questions centers on the possibility that respondents might say what they think you would like to hear. In order to satisfy yourself about the validity of the answers, frame your questions in such a way

that there is not a clear "right" answer. For example, don't merely ask, "Are you happy with the outcome of the activity?"

Planning for the collection of data. This is really a part of scheduling and staffing, but it is the most sensitive and deserves more detailed consideration than other parts of the evaluation. Two weeks "hesitation" in producing the final evaluation report may be less damaging than ten minutes hesitation in data collection. Know what each piece of information will look like when gathered, have an easy way of labeling it to identify its source, and have a clearly marked place to put the recorded information. Then plan to double-check the receipt of information, if at all possible, while you still have access to the sources of the data. If part of your information gathering is by interview or performance observation, provide discrete checklists of alternatives and categories of behavior to reduce the ambiguity of the data recorded. Finally, try out the procedures ahead of time and eliminate as many problems as possible before the actual data collection time approaches.

A DATA ANALYSIS PLAN SHOULD BE AS SIMPLE AS POSSIBLE. COLOR IT SIMPLE!

Preparing procedures to use the data. Your "data analysis plan" should be as simple as possible. Start with the decision-rationales you are to test and the information needs you identified for testing them. Your aim should be to approximate as closely as possible a statement such as "Thus, the results of the evaluation strongly support (do not support) this

argument." Presentation of tables and statistical comparisons of scores should aim to provide the foundations for these statements. Tables and comparisons that are possible to present but unrelated to any decisions should be deleted (possibly after you have satisfied your curiosity about them).

Finally, be aware of the attacks that will be made on your interpretation of the data and address them as well as possible. Take every opportunity for obtaining critiques of your report before it is released, and respond to the criticisms.

This five-step evaluation planning procedure is designed to counteract the arguments that evaluation is irrelevant and of little utility. In a small evaluation setting you cannot go through all five steps at great length--the cost might be greater than the cost of the staff development activity itself and certainly not warranted by the value of any aids to decision it would produce. But you can go through each step quickly, rather than completely ignore, for example, the identification of decision-makers. Determine the need for your evaluation and the budget feasible for it. Then you can decide how much effort to put into the evaluation. If you find that all important decision-rationales have already been adequately tested, argue against evaluation! An evaluation that is merely data collection, unrelated to decision-making, has little to recommend it.

APPENDIX A

Using Correlations for Validation of Measures

One method for validation of a paper-and-pencil measure, A, using a more extensive performance measure, B, is by calculating the correlation between them: If Measure A is valid, then those who score high on A (relative to others) should be those who score high on B. The most widely used measure of the correlation is the Pearson Product-Moment Correlation Coefficient. We briefly present the formula for its computation here to demonstrate that estimation of validity is not too difficult.

Computation of r_{AB}, based on a sample of n individuals which should be at least 30.

---

1. Add the scores on A together. Call the sum "SA."
2. Add the scores on B together. Call the sum "SB."
3. Square each score on A and add them up. Call the sum "SAA."
4. Square each score on B and add them up. Call the sum "SBB."
5. For each individual, multiply the A and B scores together, then add up the products. Call the sum "SAB."
6. Square "SA," divide the result by n, then subtract from "SAA." Call the result "VAA."
7. Square "SB," divide the result by n, then subtract from "SBB." Call the result "VBB."
8. Multiply the "SA" times "SB," divide the result by n, then subtract from "SAB." Call the result "VAB."
9. Multiply "VAA" times "VBB" and look up the square root of the product (or use an electronic calculator). Call the result "D."
10. Divide "VAB" by "D." The result, which must be between +1 and -1, is the correlation, "r_{AB}."

$$\text{or: } r_{AB} = \frac{SAB - (SA)(SB)/n}{\sqrt{(SAA - (SA)^2/n)(SBB - (SB)^2/n)}}$$

If r_{AB} is not at least +5, then either your paper-and-pencil measure or your performance measure is in need of revision! To understand the size of a correlation, we talk about the "proportion of variance accounted for," which is equal to the square of the correlation coefficient. Thus, for example, there could be four entirely independent paper-and-pencil measures, each

correlated .5 with the performance measure ($(.5)^2$ = .25, and .25 + .25 + .25 + .25 = 1.00). Any one of them, say your measure A, accounts for only 1/4th the variance in B. A value of greater than +.7 for r_{AB}, on the other hand, is quite respectable and should give you confidence in relying on the paper-and-pencil tests as substitutes for following up all participants as they later perform in the schools.

This method of validation can be extended to validate teachers' gains in terms of the improvement in average scores of their students. In this case, however, because of the many possible confounding factors (see Figure 1), any correlation r_{AB} greater than .33 should be taken as suggesting that the staff development activity is affecting student learning and any correlation greater than .45 should be cause for general satisfaction with the use of the teacher gain score, A, as a proxy for student gains. This is provided, of course, that the validation was unbiased (i.e., any sampling of teachers or students must be random, and the student gain measure was not merely a rote translation of items used to measure teacher gains).

STAFF CAREER EDUCATION SEMINAR

Executive Summary

Once a month during the 1976-77 school year, an inservice seminar met to discuss career education concepts and practices. The seminar was open to all district staff and met on Monday evenings. A modest budget was provided to pay the expenses of individuals invited to make presentations.

The overall objective of the seminar was to increase the frequency of career education activities in the classroom. This was to be accomplished by demonstrating successful activities, by discussing the benefits to be gained through career education, and by providing information sources to facilitate activities. One part of the objective was to discuss the problems some teachers had brought up concerning career education so that a district-wide program could be adapted to the particular context of our city.

As a part of the seminar's activities, participants were asked to provide feedback that would help to answer several questions.

- Should the seminar be continued?
- If so, how can it be improved?
- If so, how can attendance be increased?
- What other types of career education staff development activities would be valuable?

In addition, participants were periodically asked to list things they had done in the classrooms during the preceding month because of the existence of the seminar. Finally, a form was developed listing twenty-five different career education activities, and by administering this form to a random sample of all the district's teachers, the frequency of these activities was assessed.

Should the Seminar be Continued?

Argument in Favor

First, of 16 participants who returned the feedback questionnaire at the last meeting, 13 of them replied "yes" and only one said "no" in response to question number 1. Second, the results of our assessment showed there was a continuing lack of career education activities in

the districts' classrooms - except among those of the seminar participants. From Table 1 we can see that among nonparticipants fewer than one activity per month was the rule, but among participants in the seminar, the average was about 2 activities per month, more among those who attended the seminar for the entire year ($t_{November} = 5.21$, $t_{May} = 7.95$).

Table 1. Frequency of Career Education Activities

	by Seminar Participants		by Others	
November	65	(36 respondents)	13	(25 respondents)
	1.8	per respondent	0.5	per respondent
	1.1	standard deviation	0.6	standard deviation
February	37	(21 respondents)		
	1.8	per respondent		
	1.0	standard deviation		
May	44	(16 respondents)	14	(25 respondents)
	2.8	per respondent	0.6	per respondent
	.9	standard deviation	0.6	standard deviation

One possible explanation of these results is that seminar participants were mainly staff who were already active career educators and that they would have conducted these career education activities even without the seminar. To test this we asked each respondent to indicate whether they had implemented each activity they checked in the preceding year, too. In Table 2, it is clear that new activities were especially attributable to the seminar. Almost no new activities were being carried out by nonparticipants, but over half of the seminar participants' career education activities were new. Moreover, if we only look at "old activities," we can see that seminar participants were not nearly so much more frequent than others in their implementation of these activities ($t_{old,\ November} = 1.57$, $t_{old,\ May} = 1.32$).

Because the participants, whose volunteered efforts place them in a strong position to subjectively evaluate the seminar, favor its continuation, and because the results of assessment of the frequency of career education activities show a substantial increase related to participation, the district's career education thrust will benefit from continuing the seminar.

Table 2. Frequencies of New and Old Career Education Activities

	by Seminar Participants		by Others	
	Old Activities	New Activities	Old Activities	New Activities
November	27 0.8 per respondent 0.7 std. deviation	38 1.1 per respondent 0.5 std. deviation	11 0.4 per respondent 0.4 std. deviation	2 .08 per respondent .08 std. deviation
February	17 0.8 per respondent 0.7 std. deviation	20 1.0 per respondent 0.4 std. deviation		
May	13 0.8 per respondent 0.6 std. deviation	31 1.9 per respondent 0.4 std. deviation	13 0.5 per respondent 0.4 std. deviation	1 .04 per respondent .04 std. deviation

Argument Against

There are two reasons for discontinuing the effort. First, although the career education coordinator spent substantial effort lining up speakers and preparing workshop activities, the attendance dropped steadily from 85 at the initial meeting to 16 at the eighth meeting. If the bulk of the increased activity was due to the efforts of those 16 who participated to the end, then at this rate it would take nearly 40 years for the 600 district staff members to be helped to infuse career education into their teaching.

Second, there is evidence in several districts in the state that the "Brandex" approach more effectively increases the frequency of career education activities, and the effort involved in implementing it is about the same as the effort required for the seminar. The "Brandex" approach relies mainly on a weekly informational newsletter, and we cannot be sure that it has the same depth of effect as the seminar does on its participants, but there is no doubt that it leads to a broader base of support for career education in the teaching staff.

Because of the low attendance rate and because of the existence of a more cost-effective approach to achieving the same objective, the districts' career education thrust will best be served by discontinuing the seminar and using the effort on an alternative approach.

Final Considerations

If the seminar is continued, how can it be improved? How can attendance be increased? According to participants the content of the seminars was excellent; the main problem was the timing of the seminar and resulting conflicts with other priorities. Participants suggested several alternatives to the one-evening-per-month schedule, including afternoon seminars involving released time from classes, taping of sessions for others to listen to, a five-day workshop presenting the same scope of topics as would be covered in the seminars, and having two or three parallel series of seminars for greater flexibility. Of these, the idea of a five-day workshop deals most directly with the problem of successive attrition. Prospective participants can more easily plan for the allocation of the whole time period than in once-a-month parcels. It also makes clearer the amount of participants' time commitment and suggests the need for compensation for this time.

What other types of career education staff development activities would be valuable? In addition to the "Brandex" approach, a number of alternatives were suggested by participants. These included promoting convention attendance, arranging meetings between educators and representatives of the business/labor/industry community, sponsoring visits to successful sites, putting together discussion groups in each school, acquiring various curriculum packages, and setting up a resource center. A comparison of the costs of several approaches and of the objectives they will be likely to achieve is contained in the full evaluation report.

Conclusion

In conclusion, the evaluation determined that the Staff Career Education Seminar was successful to some extent in increasing the district's implementation of career education but that there was a problem of ensuring continued attendance. The district's career education planners should carefully consider other alternatives next year, but some type of staff development activity is clearly needed because of the infrequency of spontaneous career education activities in the classroom.

EVALUATION OF PRODUCT DEVELOPMENT

Model # 5

Steven M. Jung

General Overview

A Career Education Product is any set of educational materials, media, or written procedures that is designed to be used repeatedly to produce specified educational outcomes for a defined audience. Texbooks, films, packages of written exercises, and audio tapes often come to mind when visualizing career education products; however, packaged staff development

programs, sets of tests and evaluation instruments, and even evaluation training handbooks can be products too. The common factors are: (1) embodiment as written materials or media; (2) a set of prespecified outcomes; (3) a defined audience; and (4) intent to replicate the outcomes on several occasions, using the same materials or media.

Because the career education concept encompasses so many different audiences for whom so many diverse outcomes might be desired, many local education agencies and even individual schools or teachers have found it

necessary to develop their own career education products, as opposed to purchasing products from commercial publishers or adopting products developed elsewhere. For this reason, it is crucial that they become aware of the relatively simple and straightforward set of evaluation concepts and procedures that have come to be associated with well done product development.

These evaluation concepts and procedures, like the scientific method on which they are based, are nothing more than (1) a set of tools to help improve the effectiveness and efficiency by which desired outcomes are achieved and (2) a set of safeguards to keep developers from fooling themselves and others as to the extent of this effectiveness and efficiency. For ease of discussion, earlier authors have termed these two functions the "formative" and "summative" purposes of evaluation, and those terms will be used in this model. Formative evaluation is carried out to help improve an educational product and summative evaluation is carried out to demonstrate the overall worth of the product.

Perhaps the most important thing to remember about formative and summative evaluation is that they are closely associated with different phases in the development of a product. Formative evaluation activities are normally carried out during the conceptualization and early development of a product, before it is actually ready for "operational" use with the intended audience. Summative evaluation activities, on the other hand, are most appropriate for application to the operational version of the product, after it has undergone at least a modicum of formative trial and revision at the preimplementation phase. In fact, early attention to summative evaluation of a product to satisfy premature accountability demands may result in abandonment of a product development effort despite the fact that the product's flaws are easily correctable. There is no simple rule of thumb that dictates when summative evaluation can begin; in general, however, it should never precede the demonstration of the product's capability to produce its desired outcomes for its intended audience in small scale formative "pilot tests."

In this model, the discussion will be divided into two phases, called "preimplementation" and "implementation." It should be remembered that there is no precise dividing line between these two phases. In the first phase, a set of concepts and procedures appropriate to formative evaluation

of an educational product is described. In the second phase, concepts and procedures more appropriate for summative product evaluation will be covered.

Preimplementation Phase

Before embarking on any formative evaluation efforts, it is well to ask whether the product needs evaluating. If the nature of the product is dictated by outside constraints over which the developer has little control or

CAREER EDUCATION PRODUCTS
WAIT A MINUTE : I FEEL A PRE IMPLEMENTATION PHASE COMING ON !! THESE PRODUCTS HAVE TO BE EVALUATED !!

if time constraints are severe, and there will be little or no opportunity or time to revise the product once it has been developed, it is pointless to proceed with the steps outlined in the following paragraphs. Time and effort can be saved by moving right into the implementation phase. However, improvements can almost always be made in the effectiveness and efficiency of a product by devoting even minimal resources to formative evaluation activities; the omission of these activities in any product development effort is unwise.

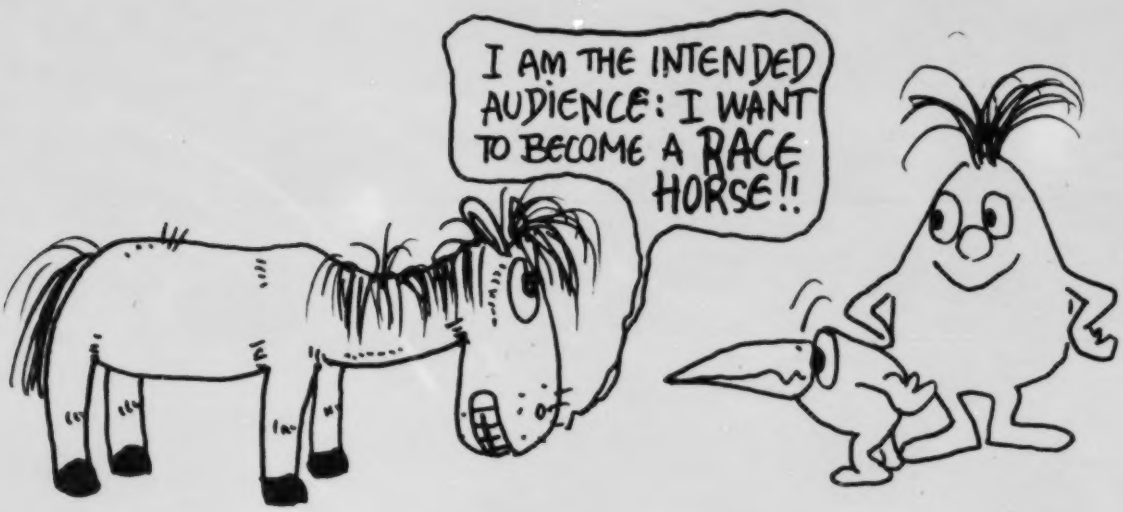
For example, in one school district a team of teachers was requested by the Assistant Superintendent for Curriculum Development to begin in March to prepare a seventh grade career education unit for infusion the coming September. Although the teachers began the task immediately and worked throughout the summer, there was no opportunity to pilot test the unit before school began. Their early implementation efforts encountered serious problems: students didn't like the learning activities they had constructed and few students were attaining the desired outcomes on end-of-unit tests.

Extensive revisions were obviously required, and implementation efforts were called off until these could be made. Because of the press of other duties, the revisions could not even begin until the following summer and an entire school year and much effort were wasted. The lesson was learned that taking shortcuts can result in more delay and waste more resources in the long run.

Review Intended Product Outcomes

It has become almost a cliché to point out the need for "good objectives" in education. Nevertheless, good objectives form the basis of any educational product and its evaluation. Before even beginning preimplementation evaluation activities, it is desirable to review the desired outcomes that are to result from application of the product. A number of questions can be asked, such as:

(1) Are the desired outcomes stated as things the intended audience will be able to do, say, feel, or think after experiencing the product? (In other words, is their attainment potentially measurable?)



(2) Is there reason to believe that the outcomes are important and not currently experienced by the intended audience? (In other words, is there a need for the product?)

(3) Are the desired outcomes feasible in view of the circumstances and appropriate for the intended audience? (In other words, is it reasonable to expect that the outcomes can be attained?)

If the answers to any of these questions appear to be no, it is essential to seek immediate revisions in the desired product outcomes. It is an axiom of evaluation that if the objectives of a product are not worth attaining, it is trivial to measure how well they are attained.

Review the Logic That Links Proposed Activities to Desired Product Outcomes

Regardless of the nature of the career education product under consideration, it will contain a set of activities to be experienced by the intended audience. This could mean passages to be read, group discussions to be led by a facilitator, slide or tape sequences to be viewed, and so forth. Although it may appear at first to be an unnecessary step, a simple logical review of the rationale that directs Activity A to the accomplishment of Outcome B may turn out to be one of the evaluation's signal accomplishments. The following questions might be asked:

(1) Do the planned activities relate to the desired outcomes? (In other words, if the planned activities are implemented, is there a reasonable chance they will produce the desired outcomes?)

(2) Are the planned activities appropriate for the intended circumstances of use with the intended audience? (In other words, are the activities practical?)

(3) Are the planned activities likely to produce unintended outcomes, as well as the intended ones, that will affect the overall impact of the product? (In other words, are there likely to be mitigating side effects?)

(4) Are the planned activities likely to be more costly in terms of resources (time, money, effort) consumed than they are worth even assuming the desired outcomes are attained? (In other words, are the activities likely to be cost-effective?)

As with the analysis of intended outcomes in the above section, "Review Intended Product Outcomes", if the answers to any of these questions point to problem areas with the product, revisions should be attempted before proceeding to empirical try-outs.

Divide the Product into Components if Necessary

If a product involves several different types of activities (for example, a student careers handbook, a staff development guide, and a videotape cassette), it is wise to divide them up and pilot test them separately. This also involves identifying distinct desired outcomes for each component and conducting a logical analysis of the rationale linking the activities of that

component to those outcomes. There is a very good reason for this practice; it has been described well by one of the nation's preeminent product developers:

> [The] point revolves around the concept of *subset optimization.* It is both reasonable and necessary to use subset criteria [in product development]. For example, although specified changes in pupil behavior represent the ultimate criterion which instructional development is attempting to optimize, it is very unwise to use this as the sole feedback basis...our staff have, at times, been very disappointed when...specific procedures such as teacher training, audiovisual segments, etc., have not improved pupil performance. One could throw out the prototype and look in a different area. But on closer examination and analysis, we have each time determined that the intended function was itself not being performed by the instructional component. That is, the teachers learned nothing or the wrong things from the training, the audiovisual segments were being "misused," etc. With subcomponents performing optimally, one has a much better likelihood of accomplishing larger functions optimally. Optimization is unlikely achievable with unreliable subcomponents. This sounds obvious, but it is almost universally overlooked in education.[1]

Conduct Limited Scale Pilot Tests with Early Versions of the Product

Early versions of a career education product can be called prototypes, first drafts, mock-ups, and so forth. Regardless of the term applied, it is

[1] Richard Schutz, "Programmatic Instructional Development," paper delivered at the 55th annual meeting of the American Educational Research Association, Minneapolis, Minnesota, March 5, 1970.

essential to subject these versions to empirical pilot tests. The pilot test represents the essence of formative evaluation, an attempt to implement the product on a microcosmic scale under highly controlled circumstances to gain as much information as possible for guiding subsequent revisions.

The most important item of information to be gained from the pilot test is whether or not the product (or, preferably, its components) are capable of producing the desired outcomes in members of the intended audience. Obviously, this requires at least three types of evaluation activities: (1) developing outcome measures; (2) identifying a small sample of representative pilot test participants and administering the outcome measures in connection with trial implementation efforts; and (3) analyzing and using the data to guide revisions. Each of these types of activities will be discussed in separate subsections.

Developing outcome measures. Outcome measures are basically an attempt to provide a public and quantitative description of the extent to which an outcome has been attained. Many volumes have been written on this subject; present readers have no doubt experienced many of them. It is not the intent of this model to offer a short course in educational and psychological measurement. There are a few basic concepts and techniques, however, that will help assure credible measures in both this phase and in the later implementation phase.

First, it is well to recognize that desired product outcomes may come in many forms. Some forms are more easily measurable than others, but this does not mean that trivial, easily-measurable outcomes should be substituted for more important, harder-to-measure outcomes. If an outcome describes something a member of the intended audience can do, say, think, or feel, it is measurable. Doing and saying represent behaviors that can be observed, recorded, elicited by tests and in conversation, obtained from existing sources (e.g., records, reports, correspondence), and so forth. Thinking and feeling are not directly observable, but they can often be inferred adequately from behaviors in a similar fashion to measurement of doing and saying. In the case of thinking and feeling outcomes, the main problem is simply adequacy of inference; in other words, "does this behavior I observe in response to a test item allow me to adequately infer the nature of the underlying thought or feeling?" The arts and sciences of psychometrics are highly organized and mathematical attempts to answer this deceptively

simple question. As early as 1951, however, John Flanagan[1] pointed out that the essence of the question was one of common sense and logical reasoning. Readers are urged to try his suggested method before all others:

(1) Describe in writing the trait to be measured, e.g., "adding a column of numbers correctly." If the trait lacks a behavioral referent (i.e., it is covert), what are some specified examples that have been used to infer its existence in the past? If it has qualitative aspects (e.g., correctly adding vs. incorrectly adding), what distinguishes high from low quality performance?[2]

(2) Analyze the behaviors representative of the trait to determine their nature, circumstances under which they are likely to occur, normal frequency of occurrence, ease with which quality judgments are rendered, and so forth.

(3) Formulate in writing a measurement "hypothesis." This hypothesis describes how the behavior might best be recorded, given the constraints of the evaluation setting. Is direct or indirect observation a possibility? Will a test or interview probably be required? If qualitative judgments will be required of responses, can they take the form of "right" or "wrong" answers or must the recorded responses be analyzed more thoroughly, e.g., by a rating scale? How precise must the estimates of the underlying trait be; for example, will an IQ or achievement test be required or simply a judgment that "this career area seems appropriate?"

A special problem of outcome measurement is the outcome that is not intended to occur immediately after exposure to the activity being evaluated. The area of career education is especially prone to such outcomes, since it is concerned with matters of school-to-work and school-to-life transition in which the effects of career education may not "mature" until long after high school graduation. Although longitudinal studies are a partial answer to this problem and will be discussed later as a part of summative evaluation, during the preimplementation phase the problem becomes a special case of inferring covert outcomes from available evidence. Again, logical analysis probably represents the best available solution. For example, if a student can describe the kinds of work carried out by members of a career cluster, can identify all of the important training and ability prerequisites for entering that cluster, has had three months of work experience in a re-

---

[1] John Flanagan, "Use of Comprehensive Rationales," Educational and Psychological Measurement, 1951, 11, 151-155.

[2] If desired outcomes represent fairly complex behaviors, e.g., making a tentative choice of a career area to explore, it may be useful to use examples again to illustrate particularly effective versus ineffective performance.

lated entry-level job, and states that s/he enjoys the work and feels s/he wants to enter that career, it is reasonable to assume that s/he will eventually attain the desired outcome of entry into a satisfying career. This caveat prompts the addition of a fourth step to the logical analysis.

(4) If any desired participant outcomes are not likely to be observable immediately after a trial implementation effort, formulate in writing an additional measurement hypothesis that identifies logical "proxy" measures which are likely to predict later attainment of the important ultimate outcomes. In other words, what are the skills, knowledges, attitudes, experiences, values, and so forth that are logically prerequisite to the attainment of the longer term outcome, and how might these be measured?

Assuming that the product developer has identified and given some prior thought to the measurement of immediate participant outcomes, the second step is to develop or adopt measures of these outcomes. Five specific techniques can probably be used to measure 95% of the important outcomes of career education:

(1) simple checklists and rating scales
(2) paper and pencil tests
(3) performance tests
(4) behavioral observations
(5) interviews

The reader's familiarity with the basic format of each of these techniques is taken for granted here.

In general, all techniques can range from more to less structured (i.e., the extent to which a participant is given the freedom to direct his/her own response; for example, a multiple choice test item is highly structured and an interview question such as "Well, what did you think of it?" is highly unstructured). More important than structure, however, are a number of considerations which reflect the adequacy of measurement. Since these considerations will be used throughout the remainder of the model, they are introduced here at some length.

• Objectivity. Will the technique yield the same score regardless of who is applying it?

It should be recognized that all measurement techniques have some deficit in objectivity. In other words, there is a continuum running from relatively objective methods to relatively subjective methods. This same

principle also applies to reliability, validity, efficiency, and nonreactivity, which will be discussed subsequently. Objective methods are those that yield similar scores no matter who is doing the scoring. Subjective methods are those that can yield widely disparate scores when applied by different persons.

In order to improve objectivity, it is necessary to establish scoring rules which enable unquivocal assignment of scores to each obtained response. For example, an "objective" paper and pencil test may require a very discrete written response, such as "True/False" or "a, b, c, or d" which, with the aid of a test key, can then be scored "right" or "wrong" by virtually everybody who might be scoring that response. A less objective essay test item requires a less discrete written response, and consequently there may be considerable disagreement among scorers as to the numerical value of that response. Objectivity can be improved for this type of item by establishing a type of key which states the general scoring rules, then lists examples of typical responses and the proper scores for each.

Another factor in improving objectivity is to remove identifying information from each response, so that a scorer will not know who wrote it. This helps control such threats to objectivity as the "halo" effect. In the "halo" effect, the responses of a person known to the scorer tend to be scaled according to the scorer's general opinion of that person rather than on the merit of the response. Information on identity can be safely removed by using code numbers rather than respondent name on all scorable documents.

All things being considered, checklists, rating scales, and paper and pencil tests are usually more objective than performance tests, observations, and interviews. During formative pilot tests, however, developers may be very interested in obtaining intuitive or idiosyncratic judgments that are not amenable to highly objective scoring. Thus, even relatively subjective techniques may have an important role at this stage if proper care is taken in interpreting their outcomes.

• Reliability. Does the technique produce data which are free from random error and thus yield a relatively "constant" score?

Reliable measurement techniques are those that yield constant scores which are relatively free from chance variation over time. For example, if one measures an object twice with a wooden ruler, the results are likely to be very consistent, or reliable. On the other hand, if a rubber band is used

to perform such multiple measurements, one may observe considerable inconsistency; the rubber band is a relatively unreliable instrument for measuring length. Lack of objectivity can lead to low reliability. Trick questions or inaccurate recording devices (e.g., unwound timers, inaccurate test keys) can also lead to low reliability.

A more important consideration in improving reliability is the way measurement techniques are applied. An application which poses the same question or situation under identical conditions to all persons being measured is generally more reliable. If possible, then, instructions and conditions should be the same for all examinees. Questions, or test items, should be phrased so that they discourage wild guessing. All examinees should be familiar with the type of response being requested. Always give a few practice items or a warm-up period so that performance will not be unduly influenced by unfamiliarity with the type of item being used. Finally, combining several items usually produces a more reliable score than a single item. For example, if one is attempting to measure knowledge of a particular concept, a more reliable estimate can be obtained by requiring satisfactory performance on four out of five items rather than on one item.

• Validity. Does the technique actually measure what it is supposed to be measuring, intuitively and by empirical demonstration, and thus yield a relatively "true" score?

For present purposes, concern is with that aspect of validity which refers to the congruence between an underlying characteristic or objective and an obtained score. For example, if a score indicates a student has mastered an objective, has s/he in fact mastered it? In this connection, one may be interested in knowing two basic facts. First, is the obtained score a measure of the desired characteristic and not some other extraneous or peripherally related characteristic? If one is interested in measuring, say, "knowledge of occupational opportunities," does the technique measure this and not reading ability, test-taking ability, etc.? Second, is the obtained score adequately representative of the entire characteristic and not some very limited aspect of the characteristic? In the example posed above, does the technique measure knowledge of a broad range of occupational opportunities or just a few limited ones? It is desirable to try to insure that measurement techniques are both closely related to and broadly representative of the outcome or characteristic being measured.

One major step in doing this is the already mentioned construction of a logical rationale for each measurement technique being utilized, to show in writing how the technique relates to the outcome or characteristic being measured. For the above example, suppose one desired student outcome for a career guidance program is for "participating students to be able to describe vocational and educational opportunities that are available to them." There are many ways to measure attainment of such an outcome; in fact, paper-and-pencil tests, performance tests, behavioral observations, and/or interviews might all be adapted to the task. Assume one decides to adopt a paper-and-pencil item such as "List at least three postsecondary educational opportunities open to everyone." The rationale for this item might look something like the following:

> The outcome is primarily cognitive, of basic knowledge type. Demonstration of attainment requires a constructed response (as opposed to a selected response or multiple-choice), with no cues or prompts. Response must show some awareness of the range of opportunities open to youth in the area. This requires scoring by persons knowledgeable about the characteristics of the area's educational opportunities. Additional scoring criteria must be designed to eliminate overlap (e.g., require "discrete" statements of opportunity) and require adequate specificity (e.g., "attend Foothill College in the general studies program," not "go to college").

The written rationale tends to: (1) force attention to important measures of the outcome (away from more trivial measures), (2) highlight practical administration and scoring details that might otherwise be overlooked, and (3) ensure that the required response is within the capability of the examinees.

The second step in helping to ensure validity is to provide sufficient measures of each important outcome or characteristic. In the above example, the designated paper-and-pencil item, while probably an acceptable one, is insufficient in itself to adequately measure attainment of the stated outcome. At the very least, another item on vocational opportunities would be required. Scores on the multiple items would then be summed to measure attainment of the outcome.

Another axiom of evaluation is that if multiple, independent measures of the same outcome produce highly similar results, one can be more confident about the validity of the measures.

Finally, measurement techniques that are relatively objective and reliable are generally more valid.

• Efficiency. Is the technique relatively cheap and easy to administer (at least within the capabilities of the person who is performing the measurement) and score?

Relatively efficient techniques are those that yield reliable and valid scores at a low cost in terms of money and examiner and examinee time. In general, this means that instruments which can be administered to groups rather than individuals, once rather than on multiple occasions, and under normal rather than contrived circumstances, are more efficient. It also means that exercises which can be scored and processed quickly (e.g., with a test key, a behavior frequency tabulator, etc.) and easily (e.g., by a clerk, a test-scoring machine, etc.) are more efficient than those which require more time (e.g., rating an essay, analyzing the content of a tape recording) and expertise (e.g., by a committee of experts). Efficiency is especially important in measuring the effects of product tryouts because the results are usually needed for immediate application in modifying the product before actual implementation.

• Non-reactivity. Does the technique unduly influence the subsequent behavior of the person to whom it is being applied?

The classic example of a highly reactive technique is uprooting a seedling daily to measure its growth. Reactivity is to be avoided unless the reactive effect of the measure is designed to be part of the activity being scrutinized, for example, when a pretest is designed to sensitize students to the material they are supposed to learn.

Relatively non-reactive measures include physical traces of past events, routinely collected records or statistics, and participant observations. An example of using physical traces as a measure is the case of a well-known research firm that made daily records of the number of cigarette butts found in ashtrays to estimate the effects of a lung cancer prevention campaign. An example of using routine records as a measure is the use of school attendance figures as a measure of the effectiveness of an anti-truancy program. Unstructured observations of such things as time to complete an activity and frequency of certain behaviors (like long disgusted sighs accompanied by throwing pencil down on desk) can often tell more than any other measure about the effectiveness of an activity. Non-reactivity is promoted when the observer is placed so as not to intrude into the situation s/he is observing.

Identifying a small sample of the intended audience and pilot testing the product. In carrying out a pilot test of a product or its components, one is most interested in (1) inferring from the results of the tryout the probable outcomes of actual implementation with the intended audience and (2) making modifications to improve the degree of outcomes attainment. Therefore, it is essential to select persons for the tryout who are at least broadly representative of the audience. In practice, this often means identifying five or

six representatives who may be predicted, because of past performance, to do exceptionally well, poorly, and average on the activity. Strict random sampling procedures in which every person in the actual target audience has an equal chance of being selected for the tryout are rarely used because of the low probability that small random samples will contain representatives of the extremes of the characteristics (ability, experience, motivation, etc.) that may cause differences in performance.

The words "evaluation design" are often mentioned in connection with arrangements of persons, an activity, and measures of effects so that inferences can be made about the probable effects of the activity on a larger group of similar persons. Although design considerations most often arise in conjunction with summative evaluation, the formative pilot test case may be used to introduce these considerations.

Design problems can quickly become very complex, especially in educational experiments that are conducted to demonstrate the relative effective-

ness of competitive products. Comparative evaluation designs, however, have no place at this particular stage of the product development process. Two so-called "pre-experimental" evaluation designs (Campbell & Stanley, 1966) are normally sufficient for the inferential needs of this stage. These are the "one group posttest only" design and the "one group pretest posttest" design. These may be diagrammed as follows, where X stands for implementation of the activity and O stands for outcome measurement.

One group posttest only design	X O
One group pretest posttest design	O_1 X O_2

The major tradeoff in selecting the latter design over the former is the value of having pre-implementation information versus the danger of producing effects solely through administration of the pretest. The latter dangers can be reduced by using nonreactive measures, which were discussed previously. Thus, the one group pretest posttest design is recommended for most tryout applications. This design will provide information on (1) the degree to which members of the target audience can already perform the behaviors referenced in the product's desired outcomes and (2) the gain in performance that is probably attributable to the product. "Probably" is used because this design does not provide safeguards to eliminate the extraneous effects of other potential causes, which are of limited concern at this point but will loom much larger later in this model.

Although the O's in the above diagrams are placed before and after the X's, this does not mean that no data are collected during the trial implementation. Product developers should be especially active in making observations and collecting other relatively subjective data to indicate where problems are occurring and why. Since the pilot test is normally conducted under conditions of developer control, ad hoc changes in materials or procedures can often be instituted and their effects immediately observed. Problems in implementation should be noted; small "glitches" in product activities under controlled circumstances can turn into major flaws under circumstances of actual implementation.

Care should be taken to ensure that obtained outcomes for each individual participant can be related to his or her pretest data and to any possibly salient personal characteristics that might have been predicted to influence the product's effectiveness (e.g., sex, ability level, motivation level, and so forth). On the basis of these findings, very sketchy evidence may be

obtained to predict whether there will be differential performance effects under actual implementation conditions.

Care should also be taken to ensure that sufficient outcome measures are applied to detect unintended outcomes, especially any that may have been predicted during the logical analysis. For example, a team of product developers at American Research Institutes (ARI) had developed a career education community leader training program with a short sound film illustrating several types of leader/group interactions. The team had carefully specified the desired outcomes of the film and constructed a pretest and posttest to measure participants' knowledge of the concepts to be taught by the film. After pilot testing a preliminary version of the film, the ARI researchers were pleased to find that most of the participants, while performing at about the chance guessing level on the pretest, got about 80% of the items right on the posttest. They were highly pleased and were formulating their plans to use the film with no further revisions when one female participant remarked on departing, "You know, that film was instructive, but the sexist language the narrator used throughout was very annoying to me!" Upon subsequent review, her objections were found to be well taken, and extensive script revisions were undertaken; without her chance remark, the film would probably have produced negative side effects in over 50% of its intended audience. It is best not to depend on chance to obtain such information. Anticipate and ask for it!

Analyzing data to guide revisions. Analyses of quantitative data for formative pilot tests need not be extensive. Percentages of items answered correctly, participants capable of performing at specified levels on structured exercises, and so forth will normally suffice. Whatever analyses are attempted should not mask the essential capabilities of the data to suggest areas where revisions are needed.

A number of decisions must be made based on the data from initial pilot tests. Some of the more common are described in this subsection. Pretest data are especially useful in helping to validate developer judgments that product activities are indeed appropriate for the intended audience. For example, if observations or paper-and-pencil test responses indicate that all selected participants can already perform satisfactorily on intended outcome measures, the need for the product is questionable. Alternatively, if the data indicate that the concepts introduced by the product are completely foreign to the intended audience, it may be necessary to add more introductory material or modify the desired outcomes to more adequately reflect the likely entry level of the audience.

Posttest data on outcome attainment may also indicate that desired outcomes are substantially beyond the capability of the product to achieve. This type of finding calls for major revisions and may even prompt a decision to rethink the need for the product. More likely these data will point to specific instructional weaknesses in various components of the product. At this point, data gathered during the pilot test become more useful in suggesting the nature of the weaknesses.

One of the most important decisions to be made is how to use the product in actual implementation. Since most products represent infusions into an ongoing instructional process, the nature of this infusion can be critical to the product's effectiveness. For example, for what types of audience might the product be most appropriate and when might it be used most appropriately? Under what circumstances is it mostly likely to be used effectively? Do there appear to be situational constraints that should be taken into account? A good preimplementation evaluation effort should set the stage for successful implementation recommendations.

Implementation Phase

Whether or not a career education product developer is accountable to some external authority for the effectiveness and efficiency of his/her product, potential users of the product need to have some assurance that it works, i.e., that it does produce the desired outcomes for the intended audience. For this reason, summative evaluation of the product's performance under operational circumstances should be carried out immediately after development has been completed. Summative evaluation activities attempt to answer such questions as these:

(1) Can the product be implemented by intended users who are not under the direct control of the developer? (In other words, is the product self-sufficient and have all the "bugs" been worked out of user instructions, activity guides, and so forth?)

(2) Is the product successful in producing its desired outcomes? Are these produced with a minimum of undesirable side effects with all members of the intended target audience?

(3) Are desired outcomes produced under a wide array of circumstances and situations? (In other words, how generalizable is the evidence of effects?)

(4) Are desired outcomes produced within applicable cost limits and likely implementation constraints? (In other words, is the product cost-effective and adaptable?)

(5) Are the outcomes stable over time? (In other words, can the effects of the product be sustained over time?)

A justifiable controversy has been waged about the appropriateness of allowing a product developer to engage in the summative evaluation of his/her product. Those who contend that summative evaluations should be conducted by "independent third parties," who have no stake in the product, argue that developers may be too biased to provide the necessary objectivity. On the other hand, proponents of developer involvement point out that no one is totally objective,[1] and that developers are more likely to have the understanding and motivation required to give their product a fair test. There is no simple solution to this controversy. The most reasonable position, and the one adopted here, suggests a compromise: developers should undertake immediate summative evaluation responsibilities, to the extent feasible, immediately after a product is made operational; if the cost and wide applicability of the product seem to justify verification and extension of the developer's results, or comparative evaluations with competing products, this should be done later by a third party.

The concepts and procedures of summative product evaluation are more detailed than those already presented for formative evaluation. But they are basically the same, involving: (1) a logical analysis; (2) development

[1] A good case in point is Brickell's (1976) description of the sad financial prognosis for the "independent" evaluator who never finds convincing evidence of success for products s/he is asked to evaluate.

of process and outcome measures; (3) collection of these measures under conditions sufficient to support empirical estimates of the effectiveness and efficiency of the product; and (4) analysis of data and preparation of reports to facilitate decisions about the overall worth of the product. These activities are discussed in the following subsections.

Analyze the Logic of the Product

As was the case prior to implementation, the first step in every evaluation of an operational product should be a thorough analysis of the logic

that directs Activity A to Outcome B in the service of some General Need C. In other fields, this is called "bench" testing, "mission," or "systems" analysis, and so forth. In educational product evaluation, it requires answers to questions such as these:

(1) Are the needs the product is designed to meet important ones?

(2) Are the needs defined in terms of required changes in the way people think, feel, or act?

(3) Do the desired outcomes of the product relate to its needs? Is there a reasonable chance that if the outcomes are attained, the needs will be met?

(4) Are the desired outcomes of the product appropriate for its intended audience and circumstances of use?

(5) Are the activities contained in the product likely to produce the desired outcomes under the intended circumstances of use?

(6) Are the activities implementable within practical time and resource constraints?

(7) Are there likely to be unintended side effects? If so, are they likely to be undesirable?

(8) Are the effects of the product measurable in the short run? If not, can suitable proxy measures be developed?

At best, negative answers to any of these questions require clarification; at the worst, they may call into question the need for subsequent evaluation steps. If the formative product evaluation was done properly, however, no negative answers should occur, and one can proceed to an empirical test of the product's logic under operational circumstances.

Develop Process and Outcome Measures

The development of outcome measures has already been discussed at some length in the preimplementation phase and will not be treated further here. It is worth mentioning, however, that the tendency to use existing standardized tests (such as tests of achievement, self-esteem, career awareness, personal and social development, and so forth) is almost always inappropriate for the summative evaluation of career education products. Outcome measures should be developed or adapted because:

(1) The desired outcomes of educational products are likely to be more focused and specific than those standardized tests are designed to measure.

(2) Conversely, the content and outcomes a standardized test is designed to cover are unlikely to match those of the product; and if there is a match, it is likely to occur for only a few items of the standardized test, with many items being irrelevant.

(3) Standardized tests are rarely suitable for the measurement of short term changes in student performance.

(4) The concept of "norms" for measures of the outcomes of a specific educational product is inappropriate.

(5) Standardized tests are unlikely to include items suitable for measuring unintended effects of the product, which are as essential in summative evaluation as they were in formative evaluation.

The term "process" measures has come to have many different and sometimes even conflicting meanings in educational evaluation. In this model, process measures refer to techniques for gathering information on the circumstances and activities associated with the actual implementation of an educational program.

At first thought, process measures may not appear to be important in the summative evaluation of an educational product, where common sense would dictate that primary attention should be paid to outcome measures. However, process measures have several important functions, including:

(1) To insure that the product was implemented as intended, so that the outcomes measured can legitimately be said to have resulted from the product and not some unknown mutation thereof.

(2) To adequately describe the circumstances and procedures of product implementation, so that readers of the summative evaluation report will know how to go about properly using the product themselves, should they decide to use it or adapt it to their own circumstances.

(3) To provide an empirical basis for future product developers to predict the effects of specified types of instructional activities on specified audiences under specified conditions.

Too many career education product evaluations in the past have been conducted without adequate attention to process measurement; this has put the product developers, potential product users, and all other educators who would engage in product development at a distinct disadvantage. The product developers don't know how to interpret the results of outcome measures in attempts to demonstrate the effects of their products. Potential users don't know how to use the product in their own circumstances so that they can obtain effects such as those demonstrated by outcome measures in the summative product evaluation. Other educators have no firm basis for predicting the effects of similar products they want to develop for use in similar circumstances. To resolve these problems, summative evaluation field tests of career education products should include attempts to measure and report at least the kinds of information discussed under the next three subheadings.

The circumstances of product utilization in the field test. The field test circumstances should be described in sufficient detail to allow others to compare them to their own circumstances. At the minimum, the following information should be recorded:

(1) How many instructional staff were involved and what were their roles?

(2) How many other types of staff were involved and what were their roles? Include here guidance and counseling personnel, administrators, parents or community volunteers, members of advisory committees, older students, and so forth.

(3) What were the essential characteristics of all persons involved? Include specialized training or credentials, if they were required, educational levels, socio-economic, racial, and female/male compositions, and so forth.

(4) What were the physical and demographic characteristics of the field test site? Include rural, suburban, or urban setting, numbers and types of buildings and classrooms, etc., used, and so forth.

(5) What components of the product were implemented, in what order, for how long?

(6) Were there any extenuating circumstances, for example, occurrence during a school holiday, use of a notable consultant, piece of equipment, or other elements not likely to be present under normal conditions of utilization?

(7) What was the "atmosphere" -- chaotic, well planned, hostile, cooperative? By whom was the field test arranged and how was it announced to the participants? Who provided the funds to pay for the field test?

(8) How active was the product developer's staff in planning and carrying out the field test? How active was the user's staff? How much advance planning was required?

(9) Under what circumstances were outcome measures collected?

Any of the common data collection techniques (e.g., checklists and rating scales, paper-and-pencil tests, performance tests, behavioral observations, and interviews) can be adapted to the requirements of process measurement, although tests are less likely to be useful, being better adapted to the needs of outcome measurement.

The costs of product utilization in the field test. One of the first questions potential product users are likely to ask is how much will it cost to implement this product in their setting, and if there are different implementation options and configurations, what are the cost trade-offs of each. Many times, the total cost of implementing a product considerably exceeds the unit cost of the product itself. Process measures collected during summative evaluation field tests should provide accurate estimates of the costs of trial implementation, using the following separate cost categories, which are likely to be meaningful to most school administrators:

(1) personnel costs of all paid staff who participated in the field test, including the units of time spent and the average cost per time unit for separate categories of staff, such as administrators, counselors, teachers, specialists, paraprofessionals, and so forth

(2) total and unit costs of equipment and materials that had to be purchased for the field test

(3) total and unit costs of building and equipment rental and/or maintenance

(4) other implementation costs, for example, publicity, substitute teachers, consultants, transportation, supplies, and so forth

(5) fixed administrative charges, if any

Potential users can divide the above into one-time and recurring costs, depending on their own plans for implementation, and can then estimate how much it would cost to implement the product initially and over its estimated life span.

The characteristics of the participants in the field test. For reasons to be discussed in the next section, it is essential that the field test participants be at least broadly representative of the intended audience for the product. They should then be described through the collection and reporting of process measures, so that potential users can judge the applicability of the field test results to their own intended audiences.

(1) How many and what types of persons (e.g., teachers, students, parents, etc.) participated?

(2) What were the characteristics of the participants? Include age, sex, grade level (if appropriate), socioeconomic status and racial composition, occupational status of parents (if appropriate), education and experience (if appropriate), and so forth.

(3) In general, what were the motivating conditions? Were the participants volunteers or conscripts? Was there an incentive for participation and/or attainment of desired outcomes? Was released time or academic credit provided, and so forth?

These and the other process data can be collected and aggregated using averages, ranges, categories showing frequencies of occurrence, and so forth. Their utilization and reporting will be discussed in a later section.

Implement a Summative Evaluation Design

As discussed earlier, an evaluation design is nothing more than an arrangement of persons, an activity, and measures of effects so that inferences can be made about the probable effects of the activity on a larger group of similar persons. In the context of a summative product evaluation, "the persons" are a representative sample of the defined audience, "the activity," is implementation of the product under user control, "the measures of effects" are intended outcome measures, and "the larger group" is the entire defined audience for the product. The inferences desired relate to the five basic summative evaluation questions listed on page 141: (1) is the product implementable under user control; (2) are desired outcomes produced, with a minimum of undesirable side effects; (3) how generalizable are the product effects under varying circumstances and with different audiences; (4) is the product cost-effective and adaptable; and (5) is outcome attainment stable over time?

The essential element of a good evaluation design is its ability to support strong inferences, especially with regard to assertions that the product as implemented caused measured effects; "strong inferences" are those that, in the definition of Platt (1964), are not easily disproved. Common assertions that can disprove an inference in a summative evaluation have been called by Campbell and Stanley (1966) "threats to validity." The purpose of implementing a good evaluation design, then, is to thwart a number of well-known threats to validly inferring that (1) the product being evaluated caused measured outcomes for the field test participants and (2) the participants are a representative sample of the entire defined audience for whom the product claims to produce intended outcomes.

Without going into a detailed enumeration of these two kinds of threats to validity, it can be pointed out that the nature of an educational product field test makes implausible a number of validity threats to the evaluation of educational programs which are larger in scale, longer in required implementation time, and less focused in implementation. Remaining are concerns such as the following.

- Could the field test participants perform well on the product's outcome measures before the product was implemented, thereby producing artificially inflated inferences about product's effects?

As mentioned earlier in the formative evaluation section, preimplementation outcome measures are almost always included to rule out this common threat. Given the fact of such a pretest, one other common threat remains.

- Were the field test participants a particularly able or well-motivated or otherwise abnormal group whose postimplementation performance on product outcome measures results in biased or nongeneralizable inferences?

This is not an easy threat to rule out. Even with a great deal of process data to show that the participants are similar to the intended audience on a number of demographic or other characteristics, it is always possible that some element inherent in their selection (e.g., a sample selected because of their high or low scores on a career knowledge test, highly motivated volunteers from a larger, less motivated population) affected their postimplementation performance on outcome measures so as to invalidate the desired inferences. About the only satisfactory way to rule out this threat completely, and the one recommended here, is to randomly create two subsamples of potential participants from the original sample of the intended audience: one of these subsamples, selected by a coin flip, participates in the product implementation and the other does not; both are then administered the same set of outcome measures. Random assignment to the participant and nonparticipant groups can be done by lottery or simply by placing all the names in alphabetical order and picking alternate names for the two subgroups. Given a situation where the field test can only accommodate a portion of the potential participants, random selection (using the 50-50 ratio or a different ratio, if there is an urgent need to accommodate more applicants or if there are places for fewer participants) is a fair and equitable way to identify participants, and a very useful one for promoting strong inferences.

When outcome measures are obtained and processed, the performance of the non-participant "comparison" group (sometimes called a "control" group) is used to provide an estimate of what the participants' scores would have been without their field test experience. It is here that "inferential statistics" (those that allow statements to be made about the probability of obtaining a difference between groups of a certain magnitude by change alone) are normally applied. The design discussed here can be diagrammed as follows:

$$R \quad O_1 \quad X \quad O_2 \qquad \text{(Participant group)}$$
$$R \quad O_3 \qquad O_4 \qquad \text{(Non-Participant group)}$$

where R stands for random assignment of persons, O stands for obtaining outcome measures, and X stands for product implementation. Campbell and Stanley (1966) call this design a randomized pretest-posttest control group design; depending on the type of outcome measures being obtained, it is appropriate to use inferential statistics like the chi-square test or t-test for independent samples to determine the "statistical significance" of obtained posttest differences between the participant and non-participant groups.[1] Formulas for calculating these standard statistical tests are contained in many basic statistics texts; they are not discussed further here.

Unfortunately, situations appropriate for the proper utilization of the above design are uncommon. There may not be enough available members of the intended audience to construct two subsamples of potential participants; all persons who might participate in the field test may be non-representative of the intended audience, posing a threat to the generalizability of the results; it may not be feasible to collect all necessary data from the non-participants, who really have no "stake" in the product or its evaluation. Under these circumstances, the product developer may have little choice but to compromise, while still maintaining the greatest possible chance of making strong inferences. For example, it may be necessary to use as a comparison basis a nonrandomly selected group of persons chosen for their willingness to cooperate in providing data. In this case, an attempt should be made to provide convincing evidence that the non-participant group is not inherently different from the participant group. The strength of inference depends directly on the strength of the evidence provided! One simple way of providing evidence is to statistically compare the pretest performance of the two groups. Although this is not entirely satisfactory by itself, because of the possibility that the pretest is not a sufficient measure of external factors that could cause differences in posttest performance not stemming from the effects of the product (such as motivation to learn), lack of pretest differences, along with descriptive evidence of the essential comparability of the groups, may be sufficient.

[1] Because of the serious reliability problems inherent in their method of calculation, simple gain scores (e.g., posttest score minus pretest score) are not recommended for use with this or any design. This problem will be discussed further in Section 3.4.

Size of groups. There may be some concern about how many participants and non-participants must be utilized in a career education product field test. The general rule is that the more participants, the greater will be the chances that inferential statistics will "detect" real differences between groups on posttest outcome measures. The generalizability of the results may also be improved. These positive aspects of the large samples must be balanced by negative aspects resulting from the following:

(1) logistical problems associated with implementing the product with large numbers of participants

(2) logistical problems associated with collecting outcome measures on a pre- and postimplementation basis for large numbers of participants and, especially, non-participants

(3) confusion of "statistically significant" differences between groups with "educationally significant" ones[1]

In general, 25-40 persons per group are probably adequate. The groups should be kept approximately the same size to avoid possible violations of the assumptions of some of the statistical tests mentioned above.

If for some reason a number of persons who take the pretest are not able to take the posttest, a new threat to validity must be tended to. While attrition may not greatly affect the statistical significance of posttest comparisons (unless the numbers of participants drop drastically below 20), there is always the possibility that characteristics of the product influenced field test participants differently, for example, by causing less able participants to drop out, leaving only the more able participants to take the posttest. In cases of non-trivial amounts of attrition from either participant or non-participant groups, the pretest scores on outcome measures for dropouts should be compared to the pretest scores of non-dropouts. The closer the scores, the less likely differential attrition is to be a legitimate threat to validity.

[1] Large groups tend to allow fairly small measured differences to be identified as statistically significant; these differences may not in fact justify the costs associated with implementing the product. In general, educational significance is a matter of judgment on the part of the evaluation user.

The special case of long term outcomes. In the previous discussion of outcome measures in the context of formation evaluation, it was pointed out that many desired career education out-comes may take a while to "mature;" for example, the desired student outcome of "finding and holding a suitable entry level job" might not be measurable until some time after high school

graduation. In that context, the best solution to measuring attainment of such outcomes was the identification of proxy measures. For summative evaluation, however, developers should consider the benefits of carrying out follow-up studies of field test participants. Such studies allow developers to (1) empirically validate their short term proxy measures by comparing immediate postimplementation data with longer term data collected to measure outcomes the proxies were intended only to predict and (2) more convincingly assert that the effects of the product are stable over time.

To conduct follow-up studies well, it is necessary to inform field test participants during the field test that further contacts are contemplated.

It is then necessary to obtain from each person an address or telephone number where s/he can be contacted. Persons for whom a further personal contact would represent a violation of privacy should be given the opportunity to opt out at this point. Follow-up questionnaires or interviews, when they are undertaken, should be especially sensitive to the need for brevity and unobtrusiveness. They should attempt to obtain only the most important data; response options should be clear and unambiguous; no extensive or otherwise time-consuming items should be included; if mailed questionnaires are used, postage paid return envelopes should be included, at the minimum.

Non-response to follow-up surveys is a special case of the differential attrition threat discussed earlier, especially when rates of return drop below 80% (as is commonly the case). As was discussed in that context, the most convincing way to rule out the possibility that respondents represent a biased subsample of the entire participant sample is to compare respondents with non-respondents on measures obtained during pre- and immediate posttesting. If significant differences are reflected in these comparisons, statistical weighting procedures must be applied to try to correct for the existing bias; these procedures are described by Wise (1977).

Analyze the Data and Prepare Reports

Data analysis considerations. Use of common inferential statistics to estimate the significance of obtained pretest and posttest differences between the participant group and the nonparticipant group has already been discussed. Several additional data analyses for summative product evaluation should be conducted routinely. These will be summarized here.

- Analyses of process-outcome linkages. Potential product users need to know the circumstances under which a career education product may be used most effectively. For this reason, at attempt should always be made to use the process data that have been collected to illustrate possible significant relationships between outcome measures and such process measures as:

  (1) distinguishing characteristics of the participant sample, especially identifiable subgroups based on socioeconomic levels, ethnic group membership, sex, past achievement levels, and so forth

  (2) distinguishing characteristics of the implementation context, such as degree of developer/user control, types of incentives in operation, order and degree of product component implementation, and so forth

The simplest way to illustrate possible relationships is to divide the posttest results on product outcome measures into two parts, one set for those participants or implementation circumstances possessing a distinguishing characteristic (or high levels of it) and another set for those participants or circumstances lacking (or with low levels of) the characteristic. Average outcome scores on other descriptive measures are then compared for the two sets, such as the illustration below:

Average Percent of Correct Responses on Knowledge of Career Options Posttest

	Participants (N = 30)	Nonparticipants (N = 30)
Girls (N = 32)	79%	48%
Boys (N = 28)	67%	23%
Overall (N = 60)	74%	38%

Several things should be noted about the above illustration in passing. First, a complete title should always accompany each data table or figure. Second, whenever data from several persons are combined, averaged, or otherwise aggregated for presentation, an indication of the number (or "N") of persons represented should always accompany the presentation. Finally, if comparisons are made based on subgroups with particularly high or low levels of some characteristic (for example, high or low achievement levels), the basis for subgroup identification should always be made clear (for example: "High (low) achievers are those scoring above (below) 50th percentile on Stanford Achievement Test, Level II Battery"). Remember that postimplementation examination of differences in subgroup performance within a randomly selected sample is acceptable and even advisable, while preimplementation division may cause serious threats to the validity of inference.

Another useful way to examine possible linkages between process and outcome measures is the application of correlation techniques. Such techniques are applicable to cases where measures are rendered in continuous or discrete scales. A continuous scale is one in which qualitative differences are represented by numerical score differences, as is normally the case with test

scores, ratings, frequency counts, and so forth. Relationships between continuous scales can be calculated using the Spearman rho or Pearson r statistics. Relationships between a continuous scale and a discrete scale or between two discrete scales can also be calculated, using other commonly available correlation methods. In any case, the demonstration of a high correlation may indicate that the product works better (or worse) under certain circumstances or with certain types of persons. This finding can be invaluable for providing sound guidance on how to utilize the product.

- Estimating reliability of measurement. The concept of measurement reliability was introduced in the section titled "Reliability" on page 133. Readers will recall that reliability relates to the constancy of a measurement technique, with high constancy being a very desirable if not essential quality for most product evaluation purposes. Reliability of measurement is so important, in fact, that many evaluation users have come to expect product evaluators to present quantitative estimates of the reliability of scores reported in an evaluation, especially for outcome measures. There are several straightforward ways to statistically estimate the reliability of a score, all based on variations of the correlation methods just discussed.

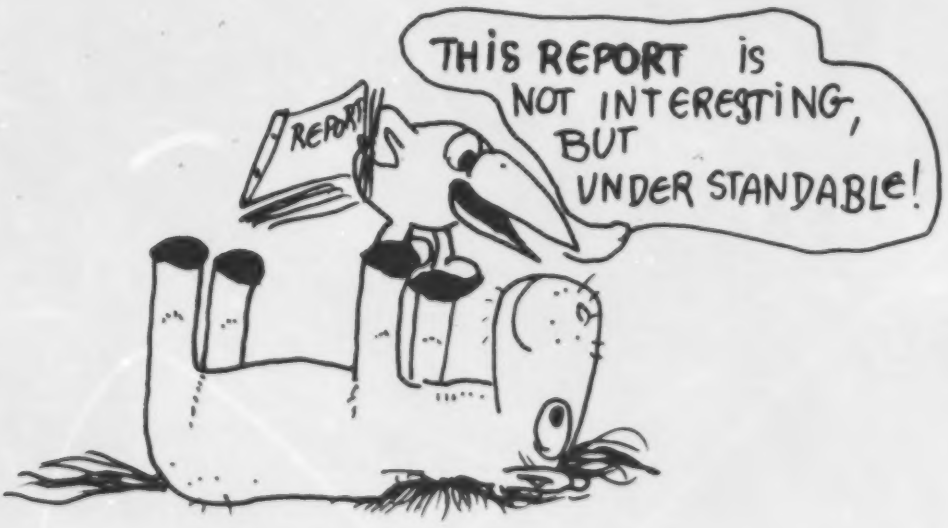
One common method, recommended here, is applicable to non-timed multi-item tests or rating scales in which all items are designed to measure the same outcome (for example, a 30-item test of knowledge about occupational training opportunities). In this so-called "split half" method, the total set of items is split into two halves (sometimes by including odd and even numbered items into the different halves), the scores on each half are calculated based on the performance of the participant and nonparticipant groups as a whole, and the two scores are correlated. The correlation is then adjusted via the Spearman-Brown formula, to take into account the fact that the full test is twice as long (and thus more reliable) than its two halves. The resultant reliability estimate should then be reported right along with the field test results. Remember that this method is not applicable to items measuring different outcomes, even if they are in the same test form. For most purposes, reliability coefficients of between .70 and .99 (1.00 is the highest possible correlation, zero is the lowest) are acceptable.

- Problems with gain scores. As previously mentioned, gain scores simply represent the difference between posttest and pretest scores for individual persons. The major statistical problem with gain scores is

their unreliability. Even though it is intuitively permissible to compare the simple gains made by the participants and nonparticipants, in fact these estimates of gains are normally only half as reliable as the pretest or posttest scores from which they are calculated. If statistical analyses of gain scores are to be undertaken, they should be supplementary to primary posttest analyses comparing the participant and nonparticipant groups.

Another common form of statistical gain analysis is to use inferential techniques such as the t-test for nonindependent samples (which has different assumptions and a different computational formula from the t-test for independent samples discussed earlier) to determine whether posttest scores are significantly better than pretest scores for the particular group. In the absence of data from a nonparticipant group, such analyses, even if they reveal highly "significant" gains, may be unconvincing, because of the threats to validity discussed earlier (for example, it may be argued that the participant group is a biased sample of the intended audience to which generalizations are desired).

Reporting summative product evaluations. It is not the purpose of this model to discuss matters of writing style, except to point out the obvious, namely that product evaluation reports should be clear and understandable for their potential readers. Rather, in this section are listed a number of things that should be included in every product evaluation report, regardless of its style or format:



(1) The desired outcomes of the product.

(2) A statement describing the intended audience, including at a minimum: entry level requirements, if any; special provisions for subgroups of especially high interest to career education, such as minorities, women, handicapped persons, veterans, etc.; and possible secondary audiences that might be impacted by the product (e.g., parents).

(3) The product's logic, including its primary components and activities and the hypotheses that direct them to the attainment of the desired product outcomes.

(4) A brief description of the pilot testing process.

(5) The field test sample and implementation context, including all relevant process data for these topics.

(6) The outcome measures used, including at a minimum: quantitative reliability estimates; a brief discussion of validity consideration; if rating scales were used, the exact wording and scoring of the ratings options (e.g., the range of possible scores and the high and low ends of the range); and sample items.

(7) All relevant results of the field test, including participant/non-participant statistical comparisons, summaries of analyses of subgroup performance, and so forth; for group comparisons on most outcome measures, figures and graphs that pictorially represent the obtained results should be strongly considered.

(8) For all results reported, information on the exact statistical analyses performed, significance levels for inferential statistics, and N's upon which comparisons were based.

(9) Information sufficient to refute any plausible threats to the validity of inference.

(10) General interpretation of the field test results, including as a minimum: the types of audiences and circumstances for which the product is most likely to be effective (e.g., process/outcomes relationships); the estimated costs of implementation in circumstances similar to the field test; unavoidable deficits in the evaluation design, if any; and needed replications or verifications of the results, if necessary.

As with the product itself, it is a good practice to have the evaluation report "pilot tested." In other words, it should be read by several representative members of the intended audience who are able to identify unclear points, missing information, erroneous logic, and so forth. Finally, the report should be proofread carefully to ensure that tables are properly labeled, attachments are properly attached, and results are reported accurately (without typographical errors).

New Educational Directions has prepared a simple checklist for reporting results of career education studies (NED, 1977) that should be useful as an aid to all product developers in assessing the adequacy of their summative evaluation reports.

FIELD TEST OF STUDENT MATERIALS TO COMBAT SEX ROLE STEREOTYPING IN CAREER EDUCATION BELLESHORE SCHOOL DISTRICT, BELLESHORE, WASHINGTON

Executive Summary

Introduction

There is ample evidence to suggest that sex role stereotyping, especially at an early age, is partially responsible for the fact that certain career patterns tend to be chosen much more frequently by members of one sex. In particular, special concern has been expressed during the past decade about the effects of stereotyping in steering more girls and women than boys and men into low paying and low status jobs. One of the often-repeated goals of career education is to ensure that a full range of educational and occupational opportunities are made available to all persons, regardless of sex. During the 1972-73 school year, staff of the Highsight School District conducted a survey that demonstrated most fifth grade students in the district held strongly stereotyped views of the occupations that were "appropriate" for women and men. As a result of this finding, the staff decided to develop a set of instructional materials for classroom use. The goals established for the materials were as follows:

- They provide content material which can expand students' perceptions of occupations open to females and to males.
- They fit within the context of subjects the teacher is already expected to cover in the classroom.
- They are self-contained.
- They are easily adapted to different classroom settings.
- They do not require any additional teaching or support staff members for implementation.

Development and pilot testing of materials by Highsight staff took place over a three year period, during which funds to augment district contributions were provided through ESEA Title III and Title IVC grants totaling almost $230,000. Three sets of materials finally evolved:

1. Occupation Simulation Packets. This set of materials consists of five separate packets, each of which may be used alone. Two packets have been developed for use with grades K-2 students, two for students in grades 3-4, and one for students in grades 5-6. These packets are designed as a hands-on career education activity for elementary students based on the "isolated skill concept." This is the singling out of a saleable work skill required for a wide variety of jobs and already possessed, in some measure, by the student. The student identifies and uses the skill in a hands-on simulated work experience. As students apply their saleable skills in a variety of job-related activities, it becomes clear that: (1) a skill is not dependent on sex; and (2) a skill required for one type of work is often transferable to another type of work. Discussion questions for teacher use with the class also emphasize these two points. Lesson plans and many required materials are included with the packets. The materials show great sensitivity to many kinds of discrimination; illustrations show a mix of races as well as sexes, and discussions of how to deal with handicapped children are included.

2. Yellow, Blue, and Red Book. This packet contains a large number of ideas for short-term activities which help the teacher and students to expand their awareness of sex-role stereotyping, and broaden their view of sex roles in the home and of appropriate job opportunities available for all qualified people. Time usage: 3-5 hours over a 3-week period. Grade levels: 1-6.

3. Many Thousand Words - Work Pictures. This looseleaf book contains a set of 8 x 10, black and white pictures of women and men, girls and boys in a variety of non-stereotyped work settings at home, school, and in the community depicting various skills and abilities. Discussion questions focus on the skills needed and on whether persons of either sex might have those skills. Time usage: 5 hours over a 3-week period. Grade levels: 1-6.

Field Test Procedures

During the 1977-78 school year, these three sets of Highsight materials were field tested by evaluation staff here in Belleshore. We attempted to determine whether the materials were effective enough to warrant our wider adoption, and to estimate their probable costs of adoption if effectiveness could be demonstrated. The field test utilized two comparable samples of students, in grades K-6, whose teachers were specifically asked to use the Highsight materials and the other whose teachers did not use the materials. Students in both samples were pre- and posttested using the "Who Should" test, developed by the University of Washington. There are two forms of this test.

K-2 form. The K-2 form has 26 questions which are read to students by the teacher. The students have only the answer sheet, not the questions. Two sample questions are: "At home, who should wash dishes? Men? Women? Or both?" "Who can use a hammer better? Boys? Girls? Or both the same?" All questions are answered by the students by circling a picture of a boy, or of a girl, or of a boy and girl standing together.

3-6 form. The grades 3-6 form of the test has four parts with a total of 41 items. As in the K-2 form, teachers read the questions aloud to students. Part I lists 19 occupations such as carpenter and nurse; students circle "man," "woman," or "both" to indicate who should do the job. Part II lists six jobs done in class such as "eraser cleaner;" response options are the same as for Part I. Parts III and IV list jobs done in the home. Response alternatives to Part III are "father," "mother," or "both," and response alternatives to Part IV are "men," "women," or "both."

Experts on sex-role stereotyping at the elementary level reviewed the tests and confirmed their content validity. Test-retest reliability was assessed in a school district that had not been exposed to instrument and materials development activities and where women's rights were not formally included anywhere in the curriculum. The test was administered twice to the same students two weeks apart. (Most Highsight materials required about two weeks of usage, so two weeks corresponded to the interval between pre- and posttests for both treatment and control classes in this study.) The test-retest correlations were .62 for the K-2 form (administered to 20 first grade students) and .88 for the 3-6 form (administered to 22 fourth grade students).

All schools participating in the field test were selected from the midrange of average schoolwide achievement test scores in the district (between the 58th to 69th percentiles on the national Fall norms of the Comprehensive Test of Basic Skills battery). Schools to receive the Highsight materials (treatment schools) were selected at random and then matched with other district schools (control schools) on the basis of location and average student achievement scores. In the treatment schools, teachers were selected at random to receive materials and participate in the pre- and posttesting; similarly, in the control schools, teachers were selected at random and asked to administer the "Who Should" test to their students twice, separated by a two week interval. All teachers received one day of training in how to administer the test, with released time and staff development credits awarded to those who actually participated. The number of teachers actually implementing the Highsight materials was 37 (out of 47 invited), while all 47 invited control group teachers elected to participate. Analysis of the characteristics of the 10 teachers who declined to participate in the field test revealed no systematic difference between them and the 37 who remained.

Field Test Results

Since treatment group students were exposed to only one of the three sets of materials, treatment group size for any particular comparison is smaller than

size of the control group, where all students were aggregated together. This was judged to pose no problem in analyzing the differences between the average treatment and control group scores, since the variability in the scores of the two groups was quite similar.

The results of the comparisons between the treatment and control groups on average (mean) pre- and posttest scores are presented in Tables 1, 2, and 3, for the Occupational Simulation Packets, the Yellow, Blue, and Red Book, and the Many Thousand Words Work Pictures respectively. The data on the next two pages show that statistically significant posttest differences were obtained, favoring the treatment groups, in 15 out of 17 comparisons. In all but 5 of these cases, pretest differences were not significant; in one of the cases where the pretest differences were significant, the difference was in favor of the control group and the direction of the difference was reversed at the posttest.

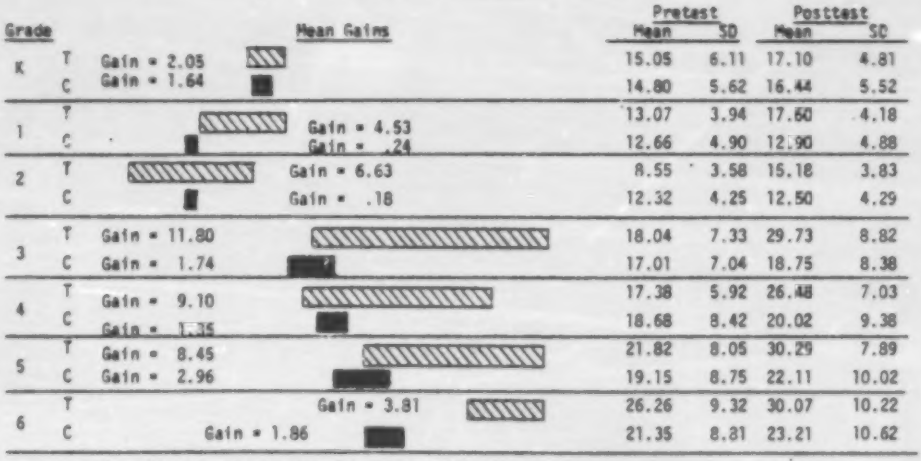
The evidence cited above demonstrates that the Highsight materials are highly effective with populations such as those in the Belleshore School District. These populations are mainly white and middle class. It is worth noting that all Highsight materials, especially those for direct use with students, are unusually attractive and present simple, intriguing activities which make use of skills most students already have regardless of their achievement levels. Furthermore, on the basis of field test costs, it is estimated that the materials can be introduced into all classrooms in the district for approximately $2280, or $.76 per student.

Conclusion

Washington State law and national legislation (e.g., Title IX of the Education Amendments of 1972) both speak to decreasing sex discrimination in all aspects of the education system. In recent studies it has been found that our schools, for the most part, are not fully demonstrating the past, present, or future participation of females in a variety of work roles in textbooks, audio-visual materials, career education programs, and teaching strategies. Across the country educators are attempting to find media and materials that present females and males in non-stereotyped occupational and personal roles. The Highsight materials fit this requirement, having been tested in an experimental atmosphere and found to be both effective and efficient.

Table 1

Occupational Simulation Packets (26 items at grades K-2, 41 items at grades 3-6)

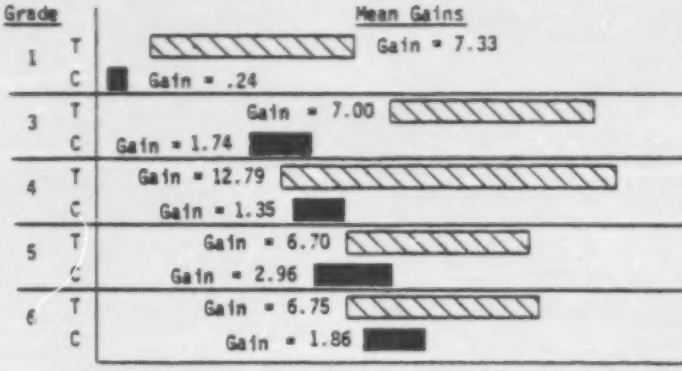


Grade		Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
K	T	15.05	6.11	17.10	4.81
	C	14.80	5.62	16.44	5.52
1	T	13.07	3.94	17.60	4.18
	C	12.66	4.90	12.90	4.88
2	T	8.55	3.58	15.18	3.83
	C	12.32	4.25	12.50	4.29
3	T	18.04	7.33	29.73	8.82
	C	17.01	7.04	18.75	8.38
4	T	17.38	5.92	26.48	7.03
	C	18.68	8.42	20.02	9.38
5	T	21.82	8.05	30.29	7.89
	C	19.15	8.75	22.11	10.02
6	T	26.26	9.32	30.07	10.22
	C	21.35	8.81	23.21	10.62

Grade	N T	N C	t-tests Pretest t	df	sig	Posttest t	df	sig
K	20	41	.15	59	NS	.48	59	NS
1	30	166	.50	194	NS	5.52	194	<.0005
2	11	156	3.33	165	<.005	2.22	165	<.025
3	49	196	.89	243	NS	7.87	243	<.0005
4	21	177	.90	196	NS	3.82	196	<.0005
5	17	194	1.30	209	NS	3.96	209	<.0005
6	27	115	2.49	140	<.05	3.12	140	<.005

Table 2

Yellow, Blue, and Red Book (26 items at Grade 1, 41 items at grades 3-6)



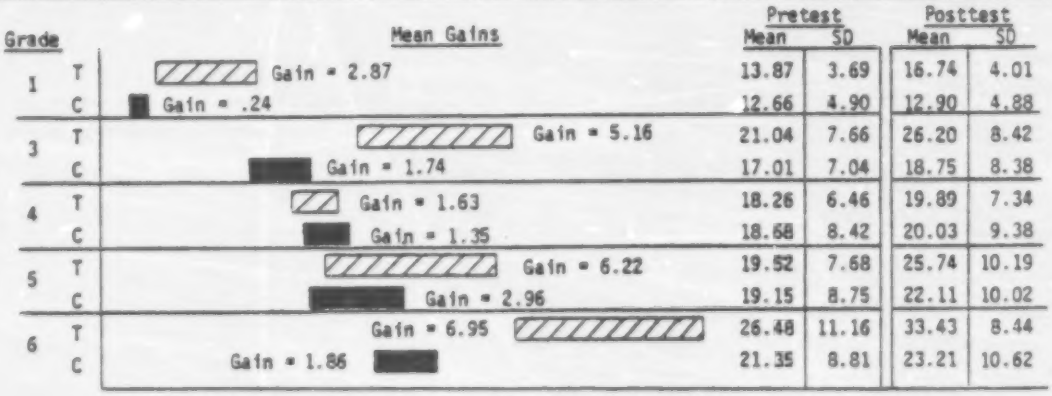
Grade		Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
1	T	13.67	6.50	21.00	4.86
	C	12.66	4.90	12.90	4.88
3	T	22.58	9.33	29.58	9.89
	C	17.01	7.04	18.75	8.38
4	T	18.09	6.83	30.88	9.29
	C	18.68	8.42	20.02	9.38
5	T	20.70	7.99	27.40	12.90
	C	19.15	8.75	22.11	10.02
6	T	20.89	10.20	27.64	10.71
	C	21.35	8.81	23.21	10.62

T = Treatment Group C = Control Group

Grade	N T	N C	t-tests Pretest t	Pretest df	Pretest sig	Posttest t	Posttest df	Posttest sig
1	6	166	.38	170	NS	4.01	170	<.0005
3	26	196	2.94	220	<.01	5.34	220	<.0005
4	33	177	.44	208	NS	6.15	208	<.0005
5	20	194	.82	212	NS	1.78	212	<.05
6	47	115	.27	160	NS	2.40	160	<.01

Table 3

Many Thousand Words - Work Pictures (26 items at grade 1, 41 items at grades 3-6)



Grade		Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
1	T	13.87	3.69	16.74	4.01
	C	12.66	4.90	12.90	4.88
3	T	21.04	7.66	26.20	8.42
	C	17.01	7.04	18.75	8.38
4	T	18.26	6.46	19.89	7.34
	C	18.68	8.42	20.03	9.38
5	T	19.52	7.68	25.74	10.19
	C	19.15	8.75	22.11	10.02
6	T	26.48	11.16	33.43	8.44
	C	21.35	8.81	23.21	10.62

T = Treatment Group C = Control Group

Grade	N T	N C	t-tests Pretest t	Pretest df	Pretest sig	Posttest t	Posttest df	Posttest sig
1	38	166	1.71	202	NS	5.10	202	<.0005
3	25	196	2.50	219	<.01	4.17	219	<.0005
4	19	177	.26	194	NS	.08	194	NS
5	73	194	.34	265	NS	2.61	265	<.01
6	23	115	2.08	136	<.01	5.06	136	<.005

EVALUATION OF CURRICULUM IMPLEMENTATION

Model #6

G. Brian Jones

General Overview

In line with its well-publicized commitment to evaluate more of the district's educational programs, a year ago a local school board appointed a seven-person panel to study the progress and impact of efforts aimed at infusing career education into the district's curriculum. Hoyt[1] talks about not building a separate curriculum empire for career education. He gets excited about what he calls "weaving" or "threading" these ideas into existing courses. This approach to career education really stresses the interrelatedness of school and work.

Specifically, the Board asked this panel to submit a report to them by the end of the school year. The support services for the study were to be provided by the district Research and Evaluation (R and E) staff. The study was to concentrate on what students were receiving in the way of career education in classrooms and what influence these activities and materials had on them.

Of considerable help to the panel during its initial planning sessions was a one-page handout summary of the general tasks the members anticipated either performing themselves or having the Research and Evaluation staff implement. Modified from a 13-step guide[2] recently published for evaluating career education programs, this overview not only helped them organize their efforts but also enabled them, the Board, and other interested persons to monitor the progress of this study. The panel's version is the 18-task statement presented as Display 1 on the next page.

The panelists agreed that the word "tasks" was much preferable to "steps," since the latter implied a rigid chronological sequence that did not seem to encourage completing a task and, on the basis of its results, returning to rework earlier tasks. While they recognized that some people react in horror to a list containing as many as 18 items, they were happy to acknowledge that in their case the more detailed their initial map, the less they would be prone to wasteful meandering during their evaluation study. We hope that you will

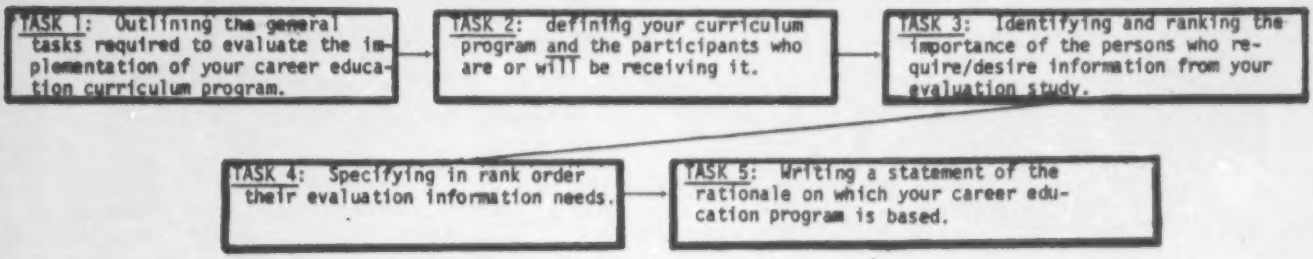
[1] Hoyt, K. B. A primer for career education. Monographs on career education. Washington, D.C.: U.S. Government Printing Office, 1977.

[2] Young, M., & Schuh, R. Evaluation and educational decision making: A functional guide to evaluating career education. Washington, D.C.: Development Associates, 1975.

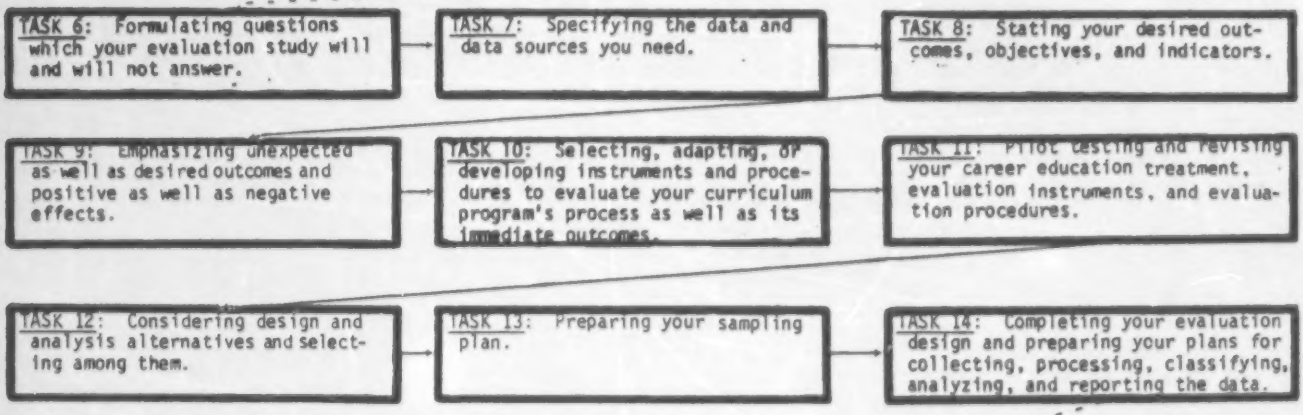
DISPLAY 1

A SUMMARY OF TASKS FOR EVALUATING THE IMPLEMENTATION OF CAREER EDUCATION CURRICULUM

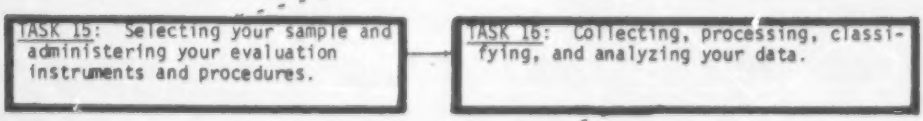
PHASE I: DEFINING AND WRITING A DESCRIPTION OF YOUR PROBLEM


TASK 1: Outlining the general tasks required to evaluate the implementation of your career education curriculum program.
TASK 2: defining your curriculum program and the participants who are or will be receiving it.
TASK 3: Identifying and ranking the importance of the persons who require/desire information from your evaluation study.
TASK 4: Specifying in rank order their evaluation information needs.
TASK 5: Writing a statement of the rationale on which your career education program is based.

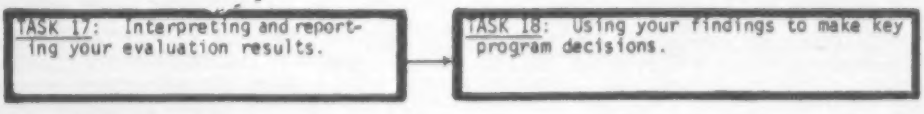
PHASE II: DEVELOPING YOUR EVALUATION DESIGN AND PLANS (AND SCHEDULE) FOR SAMPLING, COLLECTING DATA, ANALYZING DATA, AND REPORTING RESULTS


TASK 6: Formulating questions which your evaluation study will and will not answer.
TASK 7: Specifying the data and data sources you need.
TASK 8: Stating your desired outcomes, objectives, and indicators.
TASK 9: Emphasizing unexpected as well as desired outcomes and positive as well as negative effects.
TASK 10: Selecting, adapting, or developing instruments and procedures to evaluate your curriculum program's process as well as its immediate outcomes.
TASK 11: Pilot testing and revising your career education treatment, evaluation instruments, and evaluation procedures.
TASK 12: Considering design and analysis alternatives and selecting among them.
TASK 13: Preparing your sampling plan.
TASK 14: Completing your evaluation design and preparing your plans for collecting, processing, classifying, analyzing, and reporting the data.

PHASE III: IMPLEMENTING YOUR EVALUATION DESIGN AND YOUR SAMPLING AND DATA PLANS


TASK 15: Selecting your sample and administering your evaluation instruments and procedures.
TASK 16: Collecting, processing, classifying, and analyzing your data.

PHASE IV: INTERPRETING, REPORTING, AND USING YOUR EVALUATION RESULTS



find Display 1 as beneficial to your career education evaluation efforts as the panel did for their district-wide assessment. Since it is written at a level that is sufficiently general, you should be able to tailor it to your purposes.

It is now one year after the evaluation panel began its study. The panel completed its report and submitted it on-time to the district Board. In the hope that the various tasks summarized in Display 1 will become more meaningful to you, we use the remainder of this chapter to share other critical decision points the panel encountered during its drama. In each case, we briefly summarize not only the decisions they made but also the alternatives they considered and the advantages and limitations of each option. At the most critical points, we provide you with displays of their products so you can use those illustrations to expedite your own evaluation efforts. Four sections comprise the remainder of this chapter. Each section addresses one of the evaluation phases summarized in Display 1.

PHASE I: DEFINING AND WRITING A DESCRIPTION OF YOUR PROBLEM

TASK 1: Outlining the general tasks required to evaluate the implementation of your career education curriculum program.

TASK 2: defining your curriculum program and the participants who are or will be receiving it.

TASK 3: Identifying and ranking the importance of the persons who require/desire information from your evaluation study.

TASK 4: Specifying in rank order their evaluation information needs.

TASK 5: Writing a statement of the rationale on which your career education program is based.

TASK 1: Outlining the general tasks required to evaluate the implementation of your career education curriculum program.

Having agreed that the 18 tasks summarized in Display 1 adequately outlined their general approach to evaluating the implementation of career education curriculum in the district's schools, the evaluation panelists turned their attention to Task 2 -- deciding what and whom they were dealing with.

TASK 2: Defining your curriculum program and the participants who are or will be receiving it.

The panel realized that if they did not have specific information describing the career education program and participants on which their evaluation study was going to be designed, they would not have a common understanding for communicating with each other or with persons outside the panel. In addition, ambiguous information in either of these areas would make it impossible to: (1) draw conclusions about causes and effects when evaluation findings were reviewed later in the project, and (2) repeat the evaluation in order to check its procedures and results.

The panel's work on Task 2 was lessened substantially by an available report summarizing each school's career education activities and the students who participated in them during the last school year. The report also indicated which activities were planned for the current year and the anticipated types and numbers of student participants. The report, written by the R and E Director and his staff, was based on data this district's personnel had collected using a questionnaire that the Director had adapted from a 1974-75 national survey of school-based career education.[1] This report meant that the panel members did not have to gather school data for Task 2 or estimate what they thought the career education program and participants might be. The first of these other alternatives would have been expensive and time consuming, while the second would have produced results of questionable accuracy.

However, they did meet with a couple of frustrations too. The major problem was that the report did not clearly identify which career education activities were being infused into the curriculum in, or as an extension of, the classroom. As you will no doubt recall, curriculum infusion of career education was the focus of their evaluation. This dilemma was solved by one panelist who produced the classification guide presented in Display 2 on the next page. The panel used this guide to organize its interpretation of the report and to list curriculum-based career education activities, and numbers

[1] McLaughlin, D. H. Career education in the public schools, 1974-75: A national survey. Palo Alto, California: American Institutes for Research, 1976. (See Appendix B.)

DISPLAY 2

CAREER EDUCATION CURRICULUM-RELATED VARIABLES BY CATEGORY[1]

A. Types of CE Instructional Materials

1. administrative material
2. bibliography or directory
3. curriculum guide
4. evaluation instrument
5. instructional unit
6. instructional module
7. laboratory-practice manual
8. simulation description
9. student workbook or manual
10. teacher's guide
11. textbook
12. audiotape
13. cassette
14. film (8 or 16 mm)
15. filmstrip
16. record, disc
17. slide
18. tranparency
19. video tape
20. activity kit
21. game
22. puzzle
23. tools and related items
24. work folder on careers
25. radio or television program
26. community resource guide

B. Types of CE Instructional Media

1. film, filmstrip, slide, videotape
2. printed material (written word)
3. audio-visual (record, audio-tape, cassette, transparency, etc.)
4. manipulatives (puzzle, game, tools, blocks, etc.)
5. spoken word, lecture
6. visuals (bulletin board, pictures, etc.)

C. Types of CE Instructional Activities

1. self-instruction (not programmed)
2. programmed instruction
3. class discussion (group or panel)
4. individual or group research project
5. simulation or gaming
6. psychodrama
7. role play
8. field trip
9. demonstration
10. drill and practice
11. supervised work experience (paid or volunteer)
12. psychomotor activities
13. manipulative skill development
14. interview with workers
15. speaker /lecture
16. job observation /job shadowing
17. computer-assisted instruction
18. individual and group career counseling coordinated with classroom activities

D. CE Infusion Strategies

1. special career education course(s)
2. separate career education units or topics in existing courses
3. career education activities presented as enrichment options in existing courses
4. career education concepts and activities used to supplement basic subject matter in existing courses

E. Types of Organizational Patterns for CE Instruction

1. large group instruction
2. small group instruction
3. open school situation
4. unique grade organization
5. extra curricular setting
6. modular or flexible schedule
7. community organization-based classroom
8. industry-based classroom
9. business-based classroom
10. learning/interest centers in classrooms
11. classroom program closely articulated with one or more career resource centers in same building

F. Staffing Patterns for CE Instruction

1. team teaching (e.g., teachers and counselors)
2. paraprofessional support
3. peer student support
4. special support personnel
5. specially trained instructor(s)

[1]Adapted from McLaughlin, D. H. Career education in the public schools, 1974-75: A national survey. Palo Alto, California, American Institutes for Research, 1976; and Hoyt, K. B. A primer for career education. Monographs on career education. Washington, D.C.: U.S. Government Printing Office, 1977.

of students participating in them, by each grade level found in the district's schools. In defining the career education program on which the evaluation study would be designed, the panel members selected appropriate items from these six categories used in Display 2: (1) instructional materials, (2) instructional media, (3) instructional activities, (4) infusion strategies, (5) organizational patterns for instruction, and (6) staffing patterns. Those category titles were the ones the panel employed in describing their program focus and in grouping student participants by grade.

TASK 3: Identifying and ranking the importance of the persons who require/desire information from your evaluation study.

By the time the panel started to work on Task 3, it had developed a concise statement of the career education curriculum activities and the student groups on which it was going to concentrate its services. In addition, the members unanimously agreed that in order for the results of their evaluation study to be as functional as possible, they must provide information useful for decision making. But that line of thought opened up the issue of determining who the "decision makers" were. Who were the persons who would make key decisions based on the evaluation findings and conclusions? If there was more than one group of decision makers, whose decision needs would be most important, and so on?

It was easy to identify the district's Board members as the primary decision makers. They were the ones who commissioned the project and to whom the panel must report. During a lengthy discussion, the R and E Director reminded the group that some financial support for the district's career education activities stemmed from state and federal grants -- all of which had reporting requirements. Thus, two additional groups of decision makers were pinpointed -- one representing the state agency that provided financial support and the other representing the federal agency. The final group of decision makers selected by the panel was the district's teachers who were committed to infusing career education's concept and activities into the regular curriculum. To make the participation of, and reporting to, these teachers more practical, the panel chose to establish communication with three task forces which had been operating during the last three years. One task force for career education focused on each academic division. The appropriate teacher on the panel was assigned the responsibility of coordinating with each task force.

Predicting that these four levels of decision makers (school, district, state, and federal) would have many similar information needs, the panel was not too concerned about deciding which group's unique needs would take precedence over the others' needs. They could not avoid placing the Board's desires in the top priority rank nor could they eliminate or underemphasize the state or federal accountability requirements. Therefore, the needs of these two agencies were assigned a secondary status while teacher needs followed in the third rank. The panel regretfully recognized that the information needs of parents and students would not be directly addressed by the proposed evaluation study. The panel agreed that their report would highly recommend that the needs of these two groups of decision makers should be carefully considered in the next career education evaluation conducted in the district.

TASK 4: Specifying in rank order their evaluation information needs.

Having identified the decision makers who, in addition to themselves, would most influence the design of the evaluation, the panel members wanted to document and compare the top priority decision needs of each group. This was not too difficult to do for the state and federal agencies because those grant requirements that mentioned evaluation data were clearly stated in the district's contracts with both agencies. However, the same amount of specificity was not available on the Board's or the teachers' needs. If the panel proceeded without gathering more detailed information, they certainly would save money and time in the short run but later would run the risk of having these groups conclude that the evaluation findings were unresponsive to their needs.

On the other hand, panelists could assume they already knew those groups' needs or they could guess them. That would be another economical but costly option, if the results turned out to be inaccurate or so non-specific that they could not be used in the detailed planning of the evaluation study. A more solid information base seemed much more desirable. Therefore, the panel decided to interview all the Board and task force members -- a total of 25 people. They felt confident in the agreements they had made, and recorded, in Task 2 regarding the career education treatment and participants on whom this study would focus. They recognized that while collecting information from the 25 interviewees they would be able to present a coherent message of their purposes and tasks.

Each panel member agreed to conduct 3-4 interviews. A procedure using a card-sort technique was drafted by the R and E Director, a brief training session preceded the next panel meeting, interviewers were assigned, and a

schedule of interviews was adopted. In the actual data collection, each interviewer reviewed a deck of cards on which were written brief statements describing questions that the evaluation study's findings might answer. Blank cards were included so that respondents could add information needs they believed were omitted. First, the interviewees were asked to select the statements that best represented the things for which they would like answers. Then, they were encouraged to select the top five of those statements and to rank them in order of importance. Therefore, comparisons of the interviewees' responses could be made. A total of 22 interviews was conducted (two persons were unavailable and one interviewer could not complete all his interviews in time). Data were combined across all interviewees and then were compared to the state and federal grant requirements. A finalized set of evaluation information needs in rank order was printed for panel approval and subsequent use.

TASK 5: Writing a statement of the rationale on which your career education program is based.

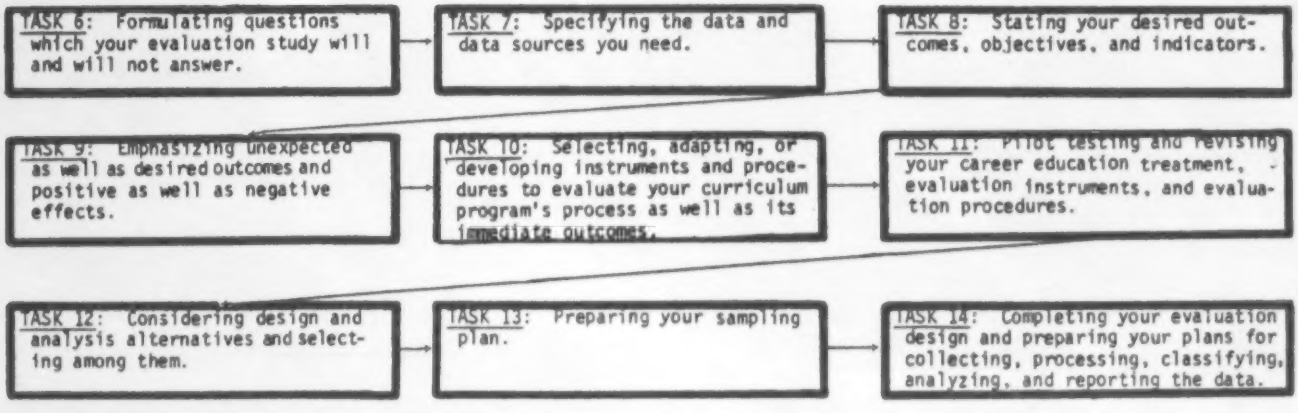
To complete their Phase I activities, the panel members combined the products of Tasks 1-4 into a "Statement of Rationale" for the career education curriculum activities on which their evaluation study would be designed. Since Task 4 had helped make some of the decision makers' hazy expectations more explicit, its results were also included in this written program summary, a document which became the fundamental blueprint for evaluation planning during the remainder of the study. The summary discussed these topics:

- the program's participants and their current needs;
- the general outcomes that these participants were expected to experience as a result of the program;
- the kinds of curriculum activities that were to be implemented in order to bring about the intended outcomes;
- a rationale which described how these general outcomes stemmed from the students' needs, and how the kinds of activities were likely to achieve the desired outcomes;
- the evaluation information needs that could be addressed by this evaluation study; and
- the general evaluation tasks the panel anticipated performing in order to conduct the remainder of this study.

The panel charged the R and E director and his staff with the responsibility of producing a first draft of the required "Statement of Rationale." The summary they produced drew heavily on two sections from the Hoyt monograph referenced earlier. The summary incorporated statements from pages 13, 27,

28, and 29 of that monograph. The panel reviewed and revised the summary's initial draft, approved the third draft, and had sufficient copies printed for dissemination during the remaining months of the evaluation study.

PHASE II: DEVELOPING YOUR EVALUATION DESIGN AND PLANS (AND SCHEDULE) FOR SAMPLING, COLLECTING DATA, ANALYZING DATA, AND REPORTING RESULTS


TASK 6: Formulating questions which your evaluation study will and will not answer.
TASK 7: Specifying the data and data sources you need.
TASK 8: Stating your desired outcomes, objectives, and indicators.
TASK 9: Emphasizing unexpected as well as desired outcomes and positive as well as negative effects.
TASK 10: Selecting, adapting, or developing instruments and procedures to evaluate your curriculum program's process as well as its immediate outcomes.
TASK 11: Pilot testing and revising your career education treatment, evaluation instruments, and evaluation procedures.
TASK 12: Considering design and analysis alternatives and selecting among them.
TASK 13: Preparing your sampling plan.
TASK 14: Completing your evaluation design and preparing your plans for collecting, processing, classifying, analyzing, and reporting the data.

TASK 6: Formulate questions which your evaluation study will and will not answer.

With the important background efforts successfully completed in its five Phase I activities, the panel was well prepared for launching into its detailed work on designing, planning, and scheduling the actual evaluation study. They began by deciding exactly what they could and could not accomplish over the next six months.

One of the panelists who had extensive interest in evaluation studies introduced an article[1] she had just read regarding the implementation of innovations. The authors emphasize that the issue of evaluating exactly how innovations are implemented is sorely neglected. They want to see evaluations give more attention to tracking and judging how a curriculum change is implemented. In fact, they are convinced that evaluations usually concentrate on how the curriculum change was planned and developed as well as what results the changes produce. The panelists agreed that this approach gives too much attention to what goes in and what comes out -- and too little to what happens between those two.

The authors focused on at least two concerns. One they call "fidelity of implementation." In short, that means the extent to which what is actually done matches how it was originally planned or intended that it would be implemented.

The authors called their second concern "mutual adaptation." That means that some people will modify the career education curriculum activities while they're implementing them. The panelists recognized that the classroom adaptations could be excellent ones -- they would just have to measure what these modifications were. They decided that if their evaluation study was heavily oriented toward process and implementation questions, such as these authors address, that did not mean their efforts would have to neglect or de-emphasize outcome measures -- assessing how students and others benefit or suffer because of career education activities related to the curriculum.

The panel went on to identify and list a total of 14 questions, grouped into these six categories:

- How do career education curriculum activities differ among the groups of students in the district and differ from the activities of students who receive little or no career education exposure?
- How do the implemented career education activities differ from those originally designed in staff training workshops?

[1]Fullan, M., & Pomfret, A. Research on curriculum and instruction implementation. Review of Educational Research, 1977, 47(1), 335-397.

- How and in what ways are career education activities modified during classroom implementation?
- In what ways do students benefit from career education activities, how do these benefits differ among the groups of students who participate, and how do they differ from outcomes experienced by students who receive little or no career education?
- How do the career education activities relate to the benefits in those situations where differences between groups are identified?
- What are the implications of this study's findings for the district's current and future career education curriculum-related activities?

In addition, the panel specified the 18 questions (e.g., How do curriculum-based career education activities influence graduates' job performance one and two years after they experience them?) that would be impossible to answer because of time and cost constraints. As the panel proceeded with its remaining tasks, it came to a fuller appreciation of the wisdom of predetermining such a comprehensive and logical set of evaluation questions. Since the chairperson sought and received Board approval of these questions, they became the "articles of agreement" for the remainder of the study. From this point on, each data effort could be, and the panel demanded that it must be, related to one or more of these questions. No further concerns related to ambiguity of study purpose were voiced.

TASK 7: Specifying the data and data sources you need.

The panel asked the R & E Director and his staff to bring to the next meeting a table that displayed for each of the 14 questions (1) the specific types of data that could answer it, and (2) the sources from which those data could be gathered (e.g., people's statements; observations of people, products, and facilities; reviews of available records and documents). The panelists recognized that if they carefully selected their data and data sources before considering the data collection instruments to be used in the study, they would be less likely to let the instruments dictate their data, as so often happens in evaluation projects.

The Director not only responded as requested but also helped the panel members by labelling each data source as primary or secondary and direct or indirect. In fact, he presented what he called a "Data Source Matrix" for which he credited the Young and Schuh evaluation guide noted earlier in this monograph.

DATA SOURCE MATRIX

Type of Source	Direct	Indirect
Primary		
Secondary		

The four terms were defined as follows:

- Primary Data Source - an original document or first-hand account or record of events.
- Secondary Data Source - material based on a primary (see above definition) or other secondary sources.
- Direct Data Source - people who are the ones described in an evaluation question.
- Indirect Data Source - people who have knowledge of the individuals described in an evaluation question.

This matrix helped panelists understand the nature of each data source as well as identify the types of sources it might be overlooking. In addition, it enabled the R & E Director to help them challenge their inclination to think that primary-direct sources should be preferred above all others. The group started to consider important issues that went far beyond asking which source has the information needed to answer each question. They discussed questions such as:

- How accessible is the required information?
- How accurate, reliable, and valid is it?
- If more than one source is available, which one is most cost-efficient to use?

Their discussion produced consensus on the data and sources for each evaluation question and made it possible for them to narrow their attention to a set of process and product objectives in Task 8.

TASK 8: Stating your desired outcomes, indicators, and objectives.

The 14 evaluation questions selected by the panel in Task 6 did not represent precise statements describing exactly what classroom teachers, and staff working with them, expected to accomplish by infusing career education activities into the curriculum. The accomplishments could be (1) actions the implementers would perform and products they would produce, or (2) ways the activities would affect the lives of the student participants and others. The panel decided to call the first set of accomplishments "process statements" and to label the second set as "outcome statements."

They even chose to enter deeper into the expanse of evaluation terminology. Now if they augmented statements in the first group with standards or "indicators" by which people's actions could be judged as desirable or undesirable, they were willing to classify these as "process objectives." Similarly, they defined "outcome objectives" as those statements that described measurable events, products or results.

The three teachers on the panel assumed responsibility for getting a set of process and outcome objectives from the career education task forces for the three academic levels. Getting statements of desired career education implementation activities and procedures was not too difficult. However, getting a coherent set of outcome objectives that described measurable student, staff, parent, and/or community outcomes of career education was not quite so easy.

In fact, these panel members spent most of their time and energy helping teachers define short-term (i.e., those that are expected immediately after a projected career education activity has occurred) and longer-term (i.e., those that are expected to develop slowly as the delayed result of a projected activity or cumulatively as a result of multiple activities) outcomes and objectives for students. They found that it was hard work, but not impossible, to help the implementers transform general statements of student outcomes (e.g., knowledge, attitudes, performance skills, or subsequent career development actions) into measurable objectives. Such objectives specified indicators that were easily and openly observable, recordable, and scorable. Such indicators had to provide uncontroversial evidence: (1) that the projected activities occurred; (2) on the degree of their occurrence, relative to what was actually planned; and (3) on the magnitude and/or quality of effects on students. Particularly instructive during these consultations and work sessions were the nine career education "learner outcomes" stated by Kenneth Hoyt[1] and others at the U.S. Office of Education (USOE). These outcome statements were presented in the introductory chapter of this Handbook.

---

[1] The outcomes are taken from: Hoyt, op cit., p. 35.
The measurement implications of these outcomes receive extensive discussion in: Hoyt, K. B. Perspectives on the problem of evaluation in career education: Monographs in career education. Washington, D.C.: U.S. Government Printing Office, 1976.
Another resource that addresses learner outcomes including career awareness, career decision making, and academic achievement is: Enderlein, T. E. A review of career education evaluation studies: Monographs on career education. Washington, D.C.: U.S. Government Printing Office, 1977.

TASK 9: Emphasizing unexpected as well as desired outcomes and positive as well as negative effects.

When he scanned the impressive array of process and outcome objectives they had developed in Task 8, one of the panelists expressed concern that if their measurement activities only addressed those statements, they would be one-sided. He arrived at two conclusions. First, such objectives describe what they should be looking for in the way of desirable career education curriculum activities, products, and outcomes but they ignore unanticipated or unintended results. He confessed that those unexpected effects, which always occur in any learning program, could be more exciting to him than could the prespecified ones. Second, he admitted he tended to look for negative effects, not just the planned positive results stated in the objectives. The panel proceeded to list these advantages of using this approach. It would enable them to:

- avoid overlooking key career education outcomes;
- use data on positive side effects of career education curriculum activities as a validity check on the achievement of desired processes and outcomes;
- use information on negative side effects to help them make decisions for improving the implementation of career education through the curriculum;
- obtain information that might help them specify new objectives; and
- have a ready supply of evidence of interest to those persons who refuse both to specify desired objectives and to accept data collected on them.

TASK 10: Selecting, adapting, or developing instruments and procedures to evaluate your curriculum program's process as well as its immediate outcomes.

The panel chairman began Task 10 with a brief discussion of a pitfall that consumes so many evaluators -- the tendency to seize available instruments and measurement procedures, inject them into what potentially was a well-designed investigation, and bestow upon those measures and the data they collect the authority of determining the objectives of the program that is being evaluated. A rather impressive statement -- obviously filled with conviction! Nor did he stop at that. He cautioned the panel members that this tendency could easily encapsulate them. To prove his point, he dis-

played six documents[1] which represent very excellent commentaries on, and compendia of, career education instruments. The group acknowledged that it would be just too inviting and simple to select measures from such "catalogs" and to forget about the hard work of either adapting them to the fine set of process and outcome objectives they had produced in Task 9 or developing their own. They were convinced that even if they had no objectives, the "select" alternative could be undesirable.

With no little hesitancy about launching into the "adopt" or "develop" options, the panel set out to review the instrument possibilities in light of their Task 7 data requirements, their Task 8 objectives, and their Task 9 resolution to employ a balanced data approach that accentuated evidence on positive and negative side effects as well as evidence on the achievement of desired effects. During the remainder of this session, they formalized the following standards they wanted their evaluation measures to meet, then they assessed the available possibilities using these standards:

- Must be compatible with the district's process and outcome objectives for the implementation of career education through the curriculum.
- Must provide appropriate coverage of program side effects as well as desired outcomes.
- Must be capable of being administered by persons and procedures that are objective, i.e., that avoid bias.

---

[1] These documents included:

(a) The Center for Vocational Education. Career education measures: A compendium of evaluation instruments. Columbus, Ohio: Ohio State University, 1976.

(b) Instructional Objectives Exchange. Illustrative criterion-referenced test specifications: Career education: Grades K-12. Los Angeles, California: Author, 1976.

(c) Hoyt, Perspectives.
NOTE: This monograph discusses measurement alternatives and problems for each of the nine learner outcomes listed in Chapter 1 of this handbook.

(d) Young & Schuh, op cit.
NOTE: This guide includes an appendix that reviews 10 recommended instruments and 4 that are considered "promising" following the authors' investigation of over 100 measures.

(e) National Assessment of Educational Progress. Career and occupational development. Denver, Colorado: Education Commission of the States, 1977.
NOTE: This resource contains 51 test items for 60 objectives related to 7 of the 9 learner outcomes noted in Chapter 1. The items are available for use with 9-, 13-, and 17-year-olds, as well as adults.

(f) Texas Partners in Career Education. Career education learning system. Arlington, Texas: Texas Education Agency, 1976.
NOTE: This project developed 17 tests, including 273 items, for 63 basic learner outcomes covering 4 of the 9 learner outcomes mentioned in Chapter 1.

- If used to assess student outcomes, must be appropriate to the participants identified in Task 2 (e.g., will have a desirable level of challenge for all students).
- If used to assess student outcomes, must be possible that same or parallel forms can be used on a pre- and post-program basis.
- If used to assess student outcomes, must be accompanied by reasonable and verifiable reliability and validity data.
- Must be relatively efficient in terms of being reasonable to purchase or produce and easy to administer and score.

In between sessions, with the assistance of the R & E staff, the panel reviewed numerous examples of the following types of instruments and evaluation procedures:

- Staff reporting forms and checklists for gathering information on whether or not planned activities stated in process objectives had been completed and when.
- Standardized norm-referenced tests and specially developed criterion-referenced tests keyed to specific student outcomes.
- Observation scales for assessing and recording events and products in natural or simulated settings.
- Interview guides for individual or small group administration for obtaining opinions, recommendations, critical incidents, and feelings from respondents.
- Self-administered questionnaires serving purposes similar to the interviews.
- Content analysis checklists for summarizing information compiled through records' searches and document reviews.
- Attitude and opinion surveys for requesting respondents to rate their degree of agreement-disagreement on selected topics.

In the following session, the panel quickly decided that it would employ a set of instruments and measurement procedures in the evaluation study. This approach not only would enable the panel to collect a comprehensive body of information that would assess all its objectives as well as program side effects, but would also provide, in a few areas, multiple findings that could be compared to determine whether or not consistent support documented: (1) the achievement of selected key objectives, and (2) the occurrence of the most impressive program side effects.

The panel also promptly dispensed with the problem of identifying all measures except those related to student outcomes. For the easily resolved instruments, the R & E Director and his staff were charged with the responsibility of adapting available measures and developing draft versions of ones needed to fill the gaps. First drafts of all such instruments were requested by a deadline two weeks away.

However, the issue of student achievement measures was not as amenable to efficient resolution. The discussion bogged down on the distinction between the purposes and advantages of criterion- and norm-referenced tests. A series of short booklets developed by RMC Research Corporation[1] proved helpful. For example, one booklet outlines four guidelines for selecting a norm-referenced test, while another summarizes conditions within which criterion-referenced instruments should and should not be used. One booklet concludes that the differences between these two types of instruments have been overemphasized. It stresses that the main difference is that norm-referenced tests are more global, i.e., they have fewer items per objective.

This discussion was fruitful because through it the panel turned its attention to three student assessment instruments that have been, or are being, national normed for many of the student outcomes on which the district's career education curriculum efforts were designed. Display 4 on the next two pages summarizes the content of these three instruments. Since each of the instruments contained a comprehensive set of subtests, the panel members realized that they could select for administration only those subtests which assessed their district's objectives for the appropriate student participant groups. They chose the Assessment of Career Development (ACD) because it most closely matched the district's student outcomes as well as provided opportunity for the addition of some locally devised items. The panel admitted that the College Board set of instruments looked promising but were

---

[1]The three booklets most pertinent to this topic are:
(a) Fagan, B., & Horst, D. Selecting a norm referenced test. Mountain View, California: RMC Research Corporation, 1976.
(b) Tallmadge, G. K., & Wood, C. T. Characteristics of eight commonly used, nationally normed tests. Mountain View, California: RMC Research Corporation, 1976.
(c) Tallmadge, G. K. Criterion-referenced tests. Mountain View, California: RMC Research Corporation, 1976.

DISPLAY 4

CONTENT OUTLINES OF THREE PUBLISHED INSTRUMENTS FOR ASSESSING CAREER EDUCATION LEARNER OUTCOMES

1. Career Maturity Inventory (Published by California Test Bureau/McGraw-Hill)

This instrument contains an Attitude Scale and a Competence Test. The Attitude Scale consists of 50 attitudinal statements assessing these five variables:

- Involvement in the choice process
- Orientation toward word
- Independence in decision making
- Preference for career choice factors
- Conceptions of the choice process

The Competence Test consists of five parts measuring these variables:

- Self-appraisal (Knowing Yourself)
- Occupational information (Knowing About Jobs)
- Goal selection (Choosing A Job)
- Planning (Looking Ahead)
- Problem solving (What Should They Do? Picking the best alternative for career development problems of hypothetical persons)

2. Assessment of Career Development (Developed by the American College Testing Program and published by Houghton Mifflin)

The Assessment of Career Development contains 162 items assessing Occupational Awareness, 20 items assessing Self Awareness, 78 items assessing Career Planning and Decision Making, and 7 items requesting Reactions to Career Guidance Experiences. More specifically, the Occupational Awareness items are divided into two groups:

- Occupational knowledge (occupational characteristics and preparation requirements)
- Exploratory occupational experiences (formal and informal ones)

The Self Awareness section addresses:

- Preferred job characteristics (job values and preferences for working conditions)
- Career plans (educational plans, occupational preferences, and certainty of occupational preferences)
- Perceived needs for help with career planning

The Career Planning and Decision Making group of items consists of two parts measuring these variables:

- Career planning knowledge (of basic career development principles, of reality factors, and of the career planning process)
- Career planning involvement (seeking information, doing and experiencing, focusing information and experience resources on specific occupational preferences, and making career plans)

The Reactions section elicits students' perceptions of help received from various aspects of school career guidance programs (e.g., information resources, guidance groups, teacher-initiated activities involving subject-related occupations).

3. Career Skills Assessment Program (Developed and published by the College Entrance Examination Board)

This set of assessment instruments has concluded its final stage of development and reportedly is available for implementation nationally during the 1978-79 school year. The assessment portion of the program addresses the following six career development areas and the skills noted under them.

Self-Evaluation and Development Skills
- Understanding individual differences
- Evaluating individual characteristics and understanding test results
- Changing personal characteristics and behavior
- Locating and interpreting information about self
- Applying knowledge about self to career opportunities

Career Awareness Skills
- Relating abilities, values, needs, and experience to career choices
- Locating, evaluating, and interpreting information for career choices
- Knowing facts about career opportunities
- Finding out about educational requirements for occupations

Career Decision-Making Skills
- Defining the problem
- Establishing an action plan
- Clarifying values
- Identifying alternatives
- Discovering probable outcomes
- Eliminating alternatives systematically
- Starting action

Employment-Seeking Skills
- Anticipating job prospects
- Finding and interpreting facts and sources of information about available jobs
- Identifying appropriately written letters, résumés, and applications for potential employers
- Describing appropriate appearance and behavior as one is interviewed and evaluated for a job
- Evaluating when a specific job fits a person's needs and interests

Work Effectiveness Skills
- Identifying the responsibilities of employers and employees to each other
- Identifying ways to develop more effective work habits
- Identifying methods to achieve effective working relationships with co-workers
- Identifying methods for managing work situations to achieve personal satisfaction
- Identifying methods for giving and receiving supervision effectively
- Identifying behavior likely to lead to advancement
- Identifying appropriate behavior for changing jobs

Personal Economics Skills
- Figuring out paychecks and income tax
- Understanding personal banking procedures
- Purchasing goods and services and paying bills
- Insuring oneself and one's possessions
- Borrowing and using credit
- Investing, including the functions of the stock market
- Understanding basic economic ideas

not yet released because validation studies were still being conducted. Finally, the panel requested that locally developed criterion-referenced tests be drafted for grade levels not addressed by the ACD.

TASK 11: Pilot testing and revising your career education treatment, evaluation instruments, and evaluation procedures.

Within the next three weeks, the panel had reviewed drafts of all instruments and evaluation procedures adapted or developed specially for this study. Suggested revisions were quickly made by R & E staff members and then copies of each such measure, except the locally developed student achievement tests, were tried out, i.e., "pilot tested," with not more than 10-12 representatives of the respondent groups for which they were designed. Each measure of student outcomes was tested with at least two classes of students. The ACD was not pilot tested since it had already been used extensively in other settings and its items and administrative and scoring directions were easily understood.

None of the panelists doubted the need for the pilot tests. They recognized that before they used these measures on a large scale, they wanted to be assured that the items themselves and the methods of administration, data collection, and scoring were feasible and desirable. They knew that if these field trials identified problems, corrections could be made before the large expense and investment of time were made in the full-scale study.

In the pilot tests, the R & E staff tried to closely approximate the actual conditions under which each instrument would be administered later. All measures were modified on the basis of pilot test results. Only one, an interview questionnaire for parents, required changes that were so extensive that an additional pilot test was conducted on the revision. The panel chairman well represented the group's sentiment when he concluded that the efforts expended on Task 11 certainly made him feel much more confident in the feasibility of the evaluation study. Indeed, he added that if they were evaluating a new career education curriculum-based program that had not been tried in the "field," he would strongly recommend that it, and not just the evaluation measures, be tried out and revised before the full-scale study began.

TASK 12: Considering design and analysis alternatives and selecting among them.

Task 12 gave one of the panelists another opportunity to impart some more of the wisdom gained in that famous measurement and evaluation course he had recently completed. He volunteered to lay out some alternative evaluation designs and related data analysis procedures for the

panel's consideration. In the Task 12 work session, he drew heavily upon a report that summarized a national survey of career education evaluations conducted in 1975-76.[1]

He began his presentation by defining "evaluation design" as "a plan that arranges persons (participants), their activities, and evaluation instruments and procedures in such a way that inferences can be made about the probable impact of those activities on a larger group of similar people." He was pleasantly surprised that no one was shattered by his definition! He then introduced the following four design alternatives, the last three of which were used to categorize all studies in the national survey. For clarity, he diagrammed each alternative, using "X" to represent the implementation of a career education activity, or program of activities, and "O" for the administration of an evaluation instrument or procedure. For each design, he also provided an example statistical technique that could be used to analyze the design's data and identified the type of question its data would answer.

He prefaced his presentation by explaining that for the panel's proposed study, a discussion of these designs would apply only to the assessment of student outcomes. Other instruments such as questionnaires for, or interviews with, staff, parents, and community representatives; direct behavior observations; and records searches would be administered using a survey design without the pre- posttest and comparison group features of some of the following designs. Such instruments would be administered while the career education activities were being implemented or after they were terminated.

- One group posttest only: X O (Program students)

  Example statistical analysis technique: Calculate frequencies, percentages, or proportions of students attaining desired performance levels on criterion-referenced tests or matching the scores of appropriate national or local norm groups on norm-referenced tests.

  Answers the question: Do participants meet predetermined performance levels or preselected age or grade norms?

- One group pre- and posttest: O_1 X O_2 (Program students)

  Example statistical analysis technique: Perform simple *t*-tests for correlated samples by comparing students' pre- and posttest scores.

  Answers the question: Do participants perform differently after the career education program than they did before they experienced it?

[1] New Educational Directions. What does career education do for kids? A synthesis of 1975-76 evaluation results. Crawfordsville, Indiana: Author, 1977.

- Two group posttest only: Randomize X_1 O_1 (Program students)
  Randomize O_2 (Comparison students)

  Example statistical analysis techniques: Perform t-tests for independent groups, or one-way analysis of variance, or X^2 test, or Mann-Whitney U test, comparing students' posttest scores.

  Answers the question: After the career education program, do participants perform differently than do other students who did not experience the program's activities?

- Two group pre- and posttest: Randomize O_1 X_1 O_2 (Program students)
  Randomize O_3 O_4 (Comparison students)

  Example statistical analysis techniques: Perform an analysis of covariance using students' adjusted posttest scores as the comparison data or perform t-tests on their amount of growth scores.

  Answers the question: After the career education program, do participants perform differently than do other students who did not experience the program but who are known to be similar to the participants?

He added that the above list could be supplemented by even more sophisticated designs. For example, he noted that ones involving more than one comparison group (receiving alternate treatments) could be diagrammed. After saying he had omitted those designs because he felt they went well beyond the panel's current purposes, he referred the group to an excellent resource on evaluation designs.[1] That jumped the panel members into a discussion of the relative merits of each of the above designs and a brief review of preliminary ideas on statistically analyzing data collected by each type of study.

The panelists agreed that their strongest approach would be to try to implement the fourth design. By using a comparison group as well as pre- and posttests, they knew that they would have a convincing basis for concluding that participant performance in this study:

- was a true gain;
- resulted from the career education curriculum rather than from extraneous factors like maturation or other instructional experiences; and
- represented that which would be emitted by a larger population of students, not of some special group that was brighter or more highly motivated.

At the same time, they knew that since all schools in the district had been implementing some career education infusion activities, they would not be able to identify "comparison" students whom they could honestly identify as never having been exposed to such career education experiences.

Therefore, they decided to use the available annual reports of each school's career education attempts and current data from a self-study instrument (devised

[1] Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally and Co., 1966.

by the panel) to classify each school in the district as either high or low on career education implementation. In this way, student outcome data collected from the "high implementation" schools would be compared to those produced in the "low implementation" schools. The R & E Director assured the panel that his staff could use its computer processing capability to handle all the study's data needs and perform the statistical tests appropriate to this design.

TASK 13: Preparing your sampling plan.

All of the panelists admitted that the proposed study had neither the schedule nor the financial resources to collect data from all possible respondents for each evaluation instrument or procedure. Therefore, by the time they turned their attention to Task 13, the issue of sampling had been mentioned on at least 27 occasions. They recognized that by sampling from the respondent "populations," they could make their study more manageable. The R & E Director convinced them that if they wished to have the findings from samples of respondents represent the opinions or performances of the total populations (e.g., student, staff, parents, and community members), their samples should be selected carefully. They agreed that since they wanted to have strong support for making generalizations about the evidence they collected from small groups, they should insure that:

- their sampling procedures were recorded for anyone's review later;
- the procedures for selecting student participants for the career education activities would permit inferences concerning the effectiveness of those activities for the sample and the total population; and
- they collected and reported data (i.e., achievement test data) to indicate how the comparison students in the low implementation group were reasonably comparable to the students in the high career implementation schools.

As the next step in Task 13, they adopted the usual way of sampling by selecting random samples from each respondent population using a table of random numbers. Only because the R & E staff knew how to use it, they chose this randomization procedure over equally acceptable ones such as using a random number generator stored in a computer, flipping a coin, or shuffling a deck of cards each of which contained a candidate's name, or systematically selecting on the basis of candidates' surnames listed alphabetically. For each respondent group, they also decided to randomly select a set of "alternates" who could be used as replacements in the case of attrition from the original selections.

The panelists realized that they had to plan for three areas of randomization. First, they decided that in order to derive generalizations about the district's curriculum-based career education activities, they would use selected classrooms with the two categories of schools (high and low implementation) as the sampling unit. From each of these two school categories, they decided to

randomly select from the district's 24 schools three elementary, two junior high, and two senior high schools from which to collect data. Second, since they knew they could not monitor the career education activities of all teachers and staff in the selected schools, they chose to collect information on a 20 percent sample of each school's classes and personnel (staff involved in career education as well as ones not involved). A similar sampling of parents and community representatives was made.

Once these key decisions were made, the R & E Director and his staff drafted a sampling plan that incorporated each of the above choices. The plan was reviewed, revised, and approved at the next panel meeting. It outlined who would be contacted, how and when the three levels of sampling would be implemented, and when data would be collected from the selected respondents. The plan represented an excellent summary of how each sample would be achieved. The panelists felt confident in the sampling decisions they had made regarding sample sizes, especially after considering the "rule of thumb" noted in the Young and Schuh guide -- i.e., that for populations of 5,000 or fewer, a sample size of 5 percent or 30 respondents is considered appropriate and safe.

TASK 14: Completing your evaluation design and preparing your plans for collecting, processing, classifying, analyzing, and reporting the data.

Having documented its sampling decisions in Task 13, the panel (with continued assistance from the R & E Branch) was able to complete its work on Phase II of its systematic evaluation approach. Once again, the R & E staff demonstrated its skills by drafting a time-task-responsibility chart covering these major remaining task areas of the proposed study:

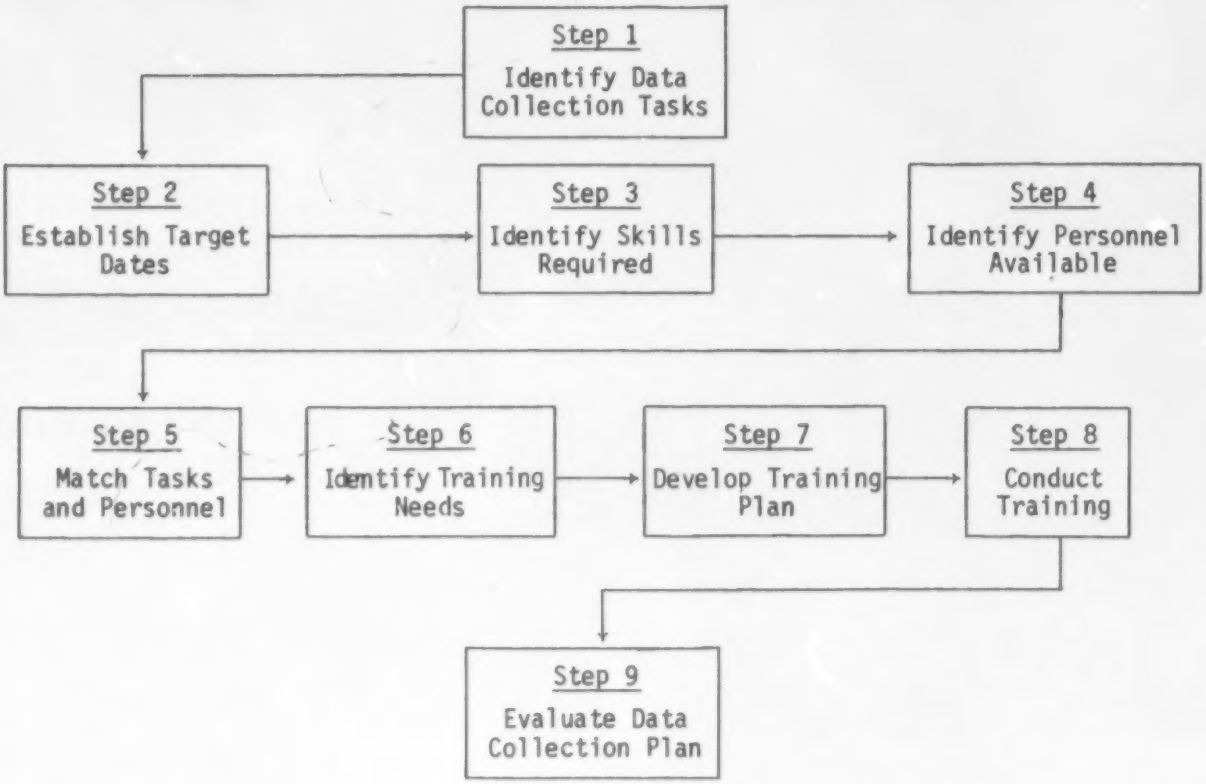
- data collection
- data processing
- data classification
- data analysis
- data reporting

Once again, the Young and Schuh guide produced a valuable tool that facilitated the R & E staff's detailed planning of the data collection efforts. Display 5 on the next page presents these authors' nine-step aid for this important task area. One of the R & E staff members adapted it to the local conditions, the project's design and schedule, and the staff's availability for data collection tasks.

The panel reviewed the draft plan produced by the R & E staff. That particular session saw the panel members making final decisions on the scope

DISPLAY 5

STEPS IN PLANNING A DATA COLLECTION EFFORT[1]

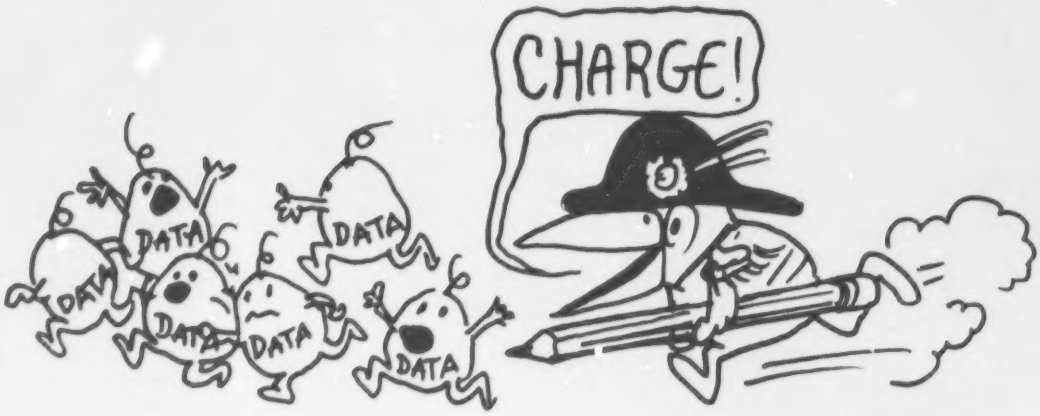


of the study based on its time and cost limitations. At this point, they also doublechecked to insure that the anticipated data would answer the evaluation questions formulated in Task 6. Now that they were able to put their hands on an approved outline of the final report that they would submit to the Board, and a plan and schedule for the report's production, they were almost tempted to feel confident that they could complete the study on schedule. Not exactly false confidence, but at least they were motivated to face the implementation problems lurking ahead of them in Phase III.

---

[1]Young & Schuh, op. cit.

PHASE III: IMPLEMENTING YOUR EVALUATION DESIGN AND YOUR SAMPLING AND DATA PLANS



TASK 15: Selecting your sample and administering your evaluation instruments and procedures. → TASK 16: Collecting, processing, classifying, and analyzing your data.

TASK 15: Selecting your sample and administering your evaluation instruments and procedures.

Under the direction of the panel chairman, the R & E staff handled the majority of the implementation activities for Tasks 15 and 16. The panelists assisted with a few chores, such as verifying that the sample of schools and respondents - within - schools selected in Task 15 indeed reflected the sampling decisions they had made in Task 13. It was during these study implementation activities that all persons connected with this project arrived at a fuller appreciation of the wisdom of having developed a complete, well-organized study plan in Task 14. This document served as a blueprint, road map, and activity program enabling them to operate more efficiently and to maintain the rigor of the evaluation design that they carefully had built into this investigation.

And it wasn't that they had no opportunity to sacrifice the study's rigor! For example, one of the randomly selected schools refused to partici-

pate; no amount of enticement or coercion changed the minds of that school's administration and staff. Rather than change the design or produce an unreasonable delay in the study's schedule, the panel ordered that one of the replacement schools, identified by the sampling procedure, be invited into the study. Happily, that school's personnel was excited about the opportunity.

Similar unexpected problems arose during the administration of the instruments and evaluation procedures the panel had selected in Task 10 and tried out in Task 11. Weaknesses in the training program for test administrators were identified and had to be quickly remediated. Staff interviews took longer than expected and scheduling binds had to be averted. Some of the parents and community representatives were unable to keep their interview appointments and the interviewers had to be encouraged, once again, to reschedule the contacts and to insure that each interview was completed.

The most serious concern was encountered in trying to monitor and record exactly what career education activities were occuring in the selected classrooms. Since in Task 6 the panel had agreed on the importance of documenting the implementation activities as well as the way teachers and students modified planned procedures while they were performing them, activity checklists had to be carefully completed by the teachers themselves. Even though these school staff members had verbalized their understanding and appreciation for the importance of this type of monitoring information, they were not perfect in their ability to record it. The classroom observations conducted by R & E evaluation personnel did not obviate the need for teachers to maintain daily recordings on their checklists. Close contacts had to be instituted with those teachers who tended to have difficulty maintaining these records. These consultations proved effective.

Similarly, the R & E staff recorded ways that the implementation of this study's design differed from the original plan. They recognized that such modifications would have major significance for later interpretation of the study's findings. Overall, the panel ended up complimenting the R & E branch for its willingness to conduct a solid evaluation study and to expend the additional energy needed to insure that the quality of the design did not suffer.

All of the panelists seemed to be gradually arriving at the conclusion that school evaluations take place in the real world, in contexts which do not permit nearly the degree of control over the conditions that is possible within a laboratory experiment. As the R & E Director pointed out, this

lack of tight control in educational treatments requires that the results of such studies be treated with less confidence than can be placed in the results of a typical laboratory experiment. On the other hand, the panel members happily argued that the real test of any treatment should be under the conditions of an operating school system, and it is in this arena that they wanted to determine the effectiveness of career education.

TASK 16: Collecting, processing, classifying, and analyzing your data.

After taking responsibility for making certain that all data were collected as close as possible to the approved schedule, the R & E evaluation team also committed to organize the data for hand and computer processing and to conduct the actual analyses selected in Task 12 and recorded in the data plan in Task 14. The following list summarizes some of the primary guidelines they used during this rather massive undertaking -- most of which was done "behind the scenes," beyond the attention of most of the panel members, except for one or two of the more curious ones who confessed to being "turned on by handling data."

- All instruments and other data records had to be checked for completeness and accuracy as soon as they were returned. Errors and omissions had to be corrected by the R & E data clerks or returned to the appropriate respondents for revision.
- Missing data records or instruments had to be identified by R & E clerks who used master lists of all respondents and measures. Contacts had to be initiated with persons responsible for collecting each school's data so that the missing measures could be found or reasons for the gaps could be documented.
- All raw data had to be classified and grouped in ways that were most appropriate for the statistical analysis methods to which they would be subjected.
- In all cases except for the ACT test results, raw data had to be coded consistent with steps outlined when each instrument and evaluation procedure was selected in Task 10. The coding process had to organize some of the data for hand tabulating by recording information on coding sheets while the remainder had to be formatted for key punching onto data cards preparatory to computer processing. The ACD tests were returned to the publisher for scoring and analysis since students' responses had been recorded on mark-sense answer

sheets and could be handled efficiently through optical scanning and entering the test scores directly on magnetic tape. The most difficult scoring and coding tasks involved the content analyses that had to be performed on respondents' answers to open-ended items contained in self-administered questionnaires and interview guides. Scorers, who completed these tasks, had to follow structured analysis and scoring rules established by the R & E Branch.

- Appropriate data analysis procedures selected in Task 12 had to be performed either manually or by the computer. Some of the procedures produced simple descriptive statistics (e.g., frequencies, percentages, means, and standard deviations) while others provided more sophisticated inferential statistics defining important relationships between different sets of raw data (e.g., estimations of the significance of obtained pretest and posttest score differences between classes of students in high career education implementation schools and classes of students in the low implementation schools). The key comparison data for this study were the class averages on the selected student outcome measures.

- Because of the unreliability of student gain or growth scores (as obtained by comparing students' pretest and posttest scores), these data had to be analyzed and studied only in the context of the more important posttest analyses contrasting the performance levels of program and comparison students. This caution had to be implemented because of the recognition that estimates of gains are usually only half as reliable as are the pretest and posttest scores from which they are calculated.

- Data summaries and displays (e.g., tables and figures) used by the R & E staff to present the findings of the manual and computer data analyses had to: (1) be labelled with a complete title that clearly described the data and the respondents; (2) specify the nature of the respondent group(s) for which data were aggregated; (3) indicate the size of the group(s); and (4) describe the types of data presented. Display 6 on the next page provides two illustrations of tabular data presentations developed by the R & E staff.

The above seven "guidelines" omit more of the activities that the R & E Branch performed for Task 16 than they include. However, these examples hit the main stages through which the data processing and analysis activities evolved. These efforts produced the data summaries and displays that were submitted to the panel members for their review and interpretation. That activity area consumed most of the panel's Task 17 time as it began Phase IV, the final two-task component of the comprehensive evaluation approach the panel had adapted to organize its entire investigation.

PHASE IV: INTERPRETING, REPORTING AND USING YOUR EVALUATION RESULTS

TASK 17: Interpreting and reporting your evaluation results.

→

TASK 18: Using your findings to make key program decisions.

DISPLAY 6

TWO EXAMPLE DATA PRESENTATIONS

TABLE 5.1

Numbers and Percentages of Grade 1-5 Students Correctly Answering 65 Percent of Items Testing Knowledge of the World of Work[1]

Types of Data / Grades	Number of Students Scoring 65 Percent or More	Percent of Students Scoring 65 Percent or More
1 (N = 213)	153	72
2 (N = 124)	97	78
3 (N = 105)	83	79
4 (N = 121)	100	83
5 (N = 95)	41	43
Totals (N = 658)	474	72

TABLE 8

Changes in Educational Plans from Pre-to-Posttesting on the Assessment of Career Development Administered to Experimental and Control Students in Grades 7-9[2]

Types of Data / Educational Plans	% of Control Students (N=181)			% of Experimental Students (N=183)		
	Pre	Post	Change	Pre	Post	Change
High School Only	13	11	-2	16	8	-8
Apprenticeship Program	2	6	+4	4	7	+3
Job Training in Military	4	2	-2	5	5	0
2 Year Community College or Technical School	11	12	+1	11	14	+3
Up to 2 Year Private Business, Trade or Technical School	6	7	+1	6	8	+2
3 Years or More of College	53	55	+2	47	51	+4
Something Else	11	7	-4	11	7	-4

[1]Adapted from: Research and Evaluation Unit. 1976-1977 final report of career education (Project EPIC). Report #14. Louisville, Kentucky: Jefferson County Public Schools, 1977.

[2]Adapted from: Gushee, T., & Organt, G. J. Evaluation of Student Progress Program of Prevocational Education Prince George's County, Maryland Public Schools. Upper Marlboro, Maryland: Prince George's County Public Schools, Career Education Department, 1976.

TASK 17: Interpreting and reporting your evaluation results.

In retrospect, all of the evaluation panel members agree that the series of meetings they devoted to reviewing the study's results; discussing their interpretations of the findings; and debating possible conclusions, implications, and recommendations from the evidence were undoubtedly the most gratifying of all their sessions. If you had observed any of their meetings, you would not have guessed that they'd ever arrive at that perception. Their "interpretation" sessions were characterized by hard work, long hours, and frequent heated discussions. The panel strove for consensus on each interpretation, with a recognition that minority opinions would be similarly documented and stated as "alternate explanations of the data" in the study's reports. One could construe this whole process as further evidence that pleasure and intense labor aren't always negatively correlated.

Another important decision they made during Task 17 was that they felt comfortable labeling some of their findings as "inconclusive." They wanted to avoid the temptation, to which they thought many evaluators succumb, of trying to derive solid conclusions from every bit of available data. Probably because the panelists enjoyed such an abundance of findings, they didn't have to be so compulsive. At the same time, they agreed that their reports would present some of their perceptions regarding "data trends" which were not strongly substantiated (e.g., findings that did not attain the panel's desired levels of statistical significance performed on data that were amenable to such comparisons).

A key part of their Task 17 activities involved making decisions about reporting the purposes, procedures, results, and interpretations of the study they had just completed. The panel discovered that the final report outline produced in Task 14 referred only to a single report. They began to discuss: (1) producing a final report consistent with the outline prepared earlier but containing no more than 60-70 pages; (2) releasing only a few copies, since not many people would want to read them, of a technical report that displayed all of their data and analyses; (3) printing at least 150 copies of an "executive summary" that condensed their final report into a readable 10-12 page booklet highlighting the purposes, objectives, procedures, findings, conclusions, and recommendations of the study; and (4) publishing about 300 copies of a two-page abstract of the investigation written in very simple terms that could be useful to parents and students as well as provide the basis for releases to the mass media.

As for report format, the panel chose to organize each of its products by the set of questions they had identified in Task 6 as being the ones for which they would provide answers. Since in Task 8 they had translated those questions into product and process objectives, they chose to maintain that orientation in their reporting. Therefore, the main body of each report would be organized around their study's objectives. They would present each objective, state where it came from and why, describe the instruments and evaluation procedures they had employed in order to see if it was achieved, summarize their results as visually as possible, and present their conclusions and recommendations related to it. All data on career education side effects would be reported either by the objective to which they most clearly relate or in a separate section.

And that's exactly how they proceeded. One resource they found helpful at this stage was a publication by a group called New Educational Directions.[1] It provides a checklist for reporting the results of career education studies. Three of the four types of final products were produced, revised, and submitted to the Board for its reactions and possible approval. In fact, the panelists even "pilot tested" each of these three reports on three or four people before they released them to the Board. This process pointed out some problem areas which the R & E staff then solved. One of the more careful readers in the R & E Branch did a fine job of proofreading each report before it was copied for the Board submission.

The technical report was not ready on time but the Board members seemed very understanding of its delay, and not too anxious to dive into it -- especially when they considered the full scope of the reports facing them. The Board is currently reviewing these documents, considering the panel's plan for disseminating them, and preparing for the culminating activity of the whole study -- Task 18, making critical decisions based on its results.

TASK 18: Using your findings to make key program decisions.

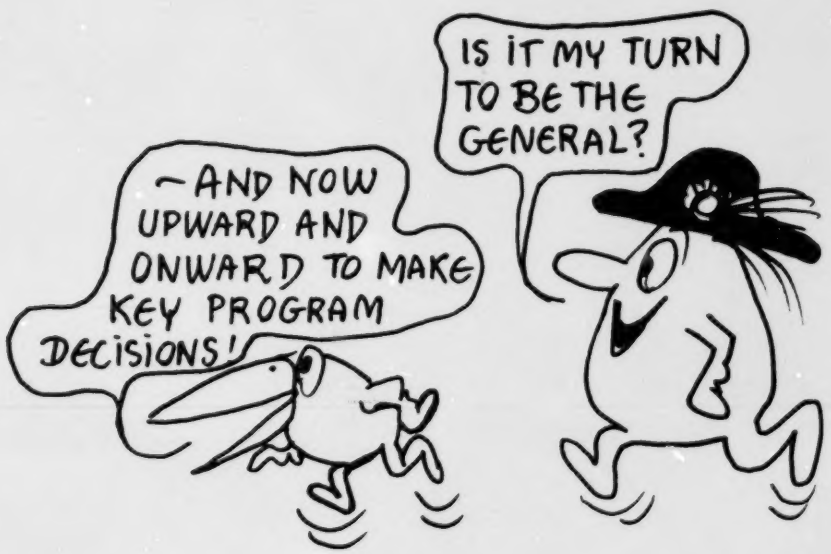
With the amount of time, energy, and pure and simple care they had devoted to this study, none of the panelists will be able to tolerate seeing the Board make glowing comments about their reports, send them individual letters of commendation, file the reports in a back shelf at the district office, and quickly return to "business as usual." The panel wants to see the Board, as the primary decision-making body for this study, arrive at choices clearly based on the

[1] New Educational Directions, Inc. Checklist for reporting results of student outcome studies in career education. Crawfordsville, Indiana: Author, 1977. (P. O. Box 307, 47933)

investigation's results and recommendations.

In fact, to encourage the decision-making process, not only of the Board but also of the other two decision-making groups (the district's teachers and state and federal representatives), the panel's reports: (1) outlined at least three decision alternatives for each of the study's major evaluation questions and objectives; (2) suggested ways that each alternative could be implemented; and (3) estimated costs for implementing each one. The alternatives dealt with possibilities such as:

- Continuing specific curriculum-based career education activities exactly as they are currently being implemented.
- Revising other activities based on the evaluation results.
- Terminating other activities because of insurmountable problems identified in the investigation.
- Developing new activities in areas identified as gaps.



- Modifying activities in order to maintain a closer compliance with state and federal government regulations.
- Attempting to replicate this study's findings to determine if results consistent with its major evidence can be substantiated or if weaknesses in its design or implementation might have confounded the results.
- Changing the resource allocations for specific career education activities in light of information regarding the achievement of process and outcome objectives relative to identifiable program costs.

By closely tying their recommendations to well-documented findings, the panelists are hopeful that the Board, and the other two groups of decision makers, will be so impressed that they will be stimulated to make at least some of the more critical decisions and to insure their implementation. The R & E Director, as panel chairperson, made an excellent presentation in submitting the study's reports to the Board. Additional debriefing sessions, including participation from other panel members, were held for the district's teachers who had participated in curriculum-based career education activities. Also attending one of these sessions were a state representative knowledgeable about grant and contract compliance requirements and a similar administrator from a regional office of the federal government. Even though they know they have achieved their mission, the panelists have not dispelled their feelings of responsibility and concern. With keen interest, they await the decision makers' actions. Meanwhile, at least four of them are willing to volunteer for the next program evaluation study.

EVALUATION OF CURRICULUM IMPLEMENTATION

Executive Summary

In line with its well-publicized commitment to evaluate more of the district's educational programs, a year ago a local school board appointed a seven-person panel to study the progress and impact of the major career education efforts implemented in the district's classrooms. The Director of Research and Evaluation was asked to chair the panel. The other six members included: (1) a parent of an elementary school student, (2) a high school sophomore, (3-5) three teachers (one from each academic division -- elementary, junior high, and senior high), and (6) the county's Coordinator for Career Education.

Much of the panel's first meeting was devoted to defining the study's purpose. The panel recognized that many career education activities were happening in the district and that it did not have to investigate all of them. Its job was to concentrate only on curriculum implementation -- on those activities taking place in the classroom. Furthermore, in concentrating on what students were receiving in the way of career education in classrooms and what influence these activities and materials had on them, the panel didn't have to evaluate why or how those curriculum changes occurred or the nature of the staff training that led to them. In short, it was supposed to narrow its attention to what career education students were receiving through the district's classes and how that was affecting them.

The remainder of that first session was devoted to clarifying the role of the evaluation panel. Of considerable help was a one-page handout the County Coordinator for Career Education had prepared at the request of the Chairperson. The panel members agreed that in order for them to play their roles in a systematic and efficient way, it would be wise to have a summary of the general tasks they anticipated either performing themselves or having the Research and Evaluation staff implement. At that point the Coordinator expressed her very favorable reaction to a 13-step guide recently published for evaluating career education programs. She recognized that many other

---

[1] Young, Malcolm, & Schuh, Russell. Evaluation and educational decision making: A functional guide to evaluating career education. Washington, D.C.: Developmental Associates, 1975.

evaluation approaches are available in the literature but she found herself pleasantly disposed to this one because it clearly addressed career education activities and it provided many details and examples. In planning for her participation in this panel, she had adapted the Young and Schuh approach in light of both her experience in career education evaluation and what she understood to be the purpose of the proposed study -- evaluating the classroom implementation of career education curriculum activities.

It was not too difficult for the other panelists to appreciate how having such an overview of their general activities would not only help them organize their efforts but also would enable them, the Board, and other interested persons to monitor the progress of this study. They agreed with the ways she had modified the Young and Schuh product and, in the true mold of any of history's great panel members worth their appointments, they made further additional modifications. This resulted in the 18-task statement presented as Display 1 on the next page.

However, there were two of her ideas with which they had no quarrel. First, they agreed that the word "tasks" was much preferable to "steps," since the latter implied a rigid chronological sequence that did not seem to encourage completing a task and, on the basis of its results, returning to rework earlier tasks. Second, they liked her lack of concern for adding 5 tasks to the original 13. While they recognized that some people react in horror to a list containing as many as 18 items, they were happy to acknowledge that in their case the more detailed their initial map, the less they would be prone to wasteful meandering during their evaluation study. We hope that you will find Display 1 as beneficial to your career education evaluation efforts as the panel did for their districtwide assessment. Since it is written at a level that is sufficiently general, you should be able to tailor it to your purposes.

This chapter shares other critical decision points the panel encountered as it performed the various tasks summarized in Display 1. In each case, the chapter briefly summarizes not only the decisions the panelists made but also the alternatives they considered and the advantages and limitations of each option. Where it best seems to communicate the panel members' deliberations, excerpts of their dialogue are presented. At the most critical points, the chapter provides displays of their products so you can use those illustrations to expedite your own evaluation efforts. Four sections comprise this chapter. Each section addresses one of the evaluation phases summarized in Display 1.

DISPLAY 1

A SUMMARY OF TASKS FOR EVALUATING THE IMPLEMENTATION OF CAREER EDUCATION CURRICULUM

PHASE I: DEFINING AND WRITING A DESCRIPTION OF YOUR PROBLEM

TASK 1: Outline the general tasks required to evaluate the implementation of your career education curriculum program.

TASK 2: Define your curriculum program and the participants who are or will be receiving it.

TASK 3: Identify and rank the importance of the persons who require/desire information from your evaluation study.

TASK 4: Specify in rank order their evaluation information needs.

TASK 5: Write a statement of the rationale on which your career education program is based.

PHASE II: DEVELOPING YOUR EVALUATION DESIGN AND PLANS (AND SCHEDULE) FOR SAMPLING, COLLECTING DATA, ANALYZING DATA, AND REPORTING RESULTS

TASK 6: Formulate questions which your evaluation study will and will not answer.

TASK 7: Specify the data and data sources you need.

TASK 8: Stating your product and process outcomes, objectives, and indicators.

TASK 9: Emphasizing unexpected as well as desired outcomes and positive as well as negative effects.

TASK 10: Selecting, adapting, or developing instruments and procedures to evaluate your curriculum program's process as well as its immediate products (outcomes).

TASK 11: Pilot testing and revising your career education treatment, evaluation instruments, and evaluation procedures.

TASK 12: Considering design and analysis alternatives and selecting among them.

TASK 13: Preparing your sampling plan.

TASK 14: Completing your evaluation design and preparing your plans for collecting, processing, classifying, analyzing, and reporting the data.

PHASE III: IMPLEMENTING YOUR EVALUATION DESIGN AND YOUR SAMPLING AND DATA PLANS

TASK 15: Selecting your sample and administering your evaluation instruments and procedures.

TASK 16: Collecting, processing, classifying, and analyzing your data.

PHASE IV: INTERPRETING, REPORTING, AND USING YOUR EVALUATION RESULTS

TASK 17: Interpreting and reporting your evaluation results.

TASK 18: Using your findings to make key program decisions.

SELECTED REFERENCES

FOR USE WITH

ALL SIX MODELS

SELECTED REFERENCES

To avoid making the models cumbersome we chose not to interrupt the flow with detailed explanations of each step. Such explanations are common across models, and would be redundant if included in each.

As you adapt one or more of the six models presented, you may need additional help in planning and/or implementing one or more steps of your evaluation. An alphabetical list of selected references which the authors have found helpful is presented here. Each reference is annotated to give you an idea of its content. Each reference also has been keyed to a set of 17 topics which represent the areas in which you may need additional help. Check the list below for the topic you wish to pursue further, and you will find the numbers of those references most likely to give you the information you need.

Topics

1. Comparison standards -2-3-15-17-30-37-73
2. Data collection, reduction, presentation -1-2-3-6-13-15-25-30-37-53-71-73
3. Evaluation decisions/questions -1-2-3-21-25-35-37-53-71-73
4. Evaluation design -2-3-14-15-17-30-37-42-47-53-71-73
5. Evaluation models -9-27-37-38-40-64-65-67-73
6. Evaluation of facilities -18-26
7. Evaluation of product development -45-52
8. Evaluation problems and how to avoid them -8-17-33-34-61
9. Instrument development -3-9-15-16-28-36-45-58-62-69
10. Instrument selection -1-3-7-13-15-16-20-22-30-31-37-43-47-53-70-71-73
11. Instruments for evaluating career education activities -4-8-16-39-42-43-48-56-63-66
12. Needs assessment -2-13-37-53-73
13. Planning and managing evaluations -10-44-51-53-68
14. Reporting procedures and formats -1-11-13-53-71-73
15. Reports of evaluated programs -5-12-19-24-29-30-32-34-49-50-55-67
16. State-of-the-art in career education -10-32-34-42-46-57-60
17. Statistical analyses -2-3-13-30-37-47-53-67-71-73

1. Adams, K. A., & Walker, J. P. Improving the accountability of career education programs: Evaluation guidelines and checklists. Columbus, Ohio: Ohio State University, Center for Vocational Education, 1977.

   Presents 13 units on evaluating career education programs. Units 1 through 8 address the planning phases: purpose, audience, questions, process, staff, responsibility/authority, uniqueness and planning standards. Unit 9 treats instruments, and units 10 through 13 treat politics, report standards, dissemination and use. A number of forms and checklists are presented to help the evaluator assess the adequacy of his/her procedures.

2. Anderson, S. B., Ball, S., Murphy, R., & Rosenthal, E. Anatomy of evaluation: Important concepts and techniques in evaluating education/training programs. Washington, D.C.: National Technical Information Service, U.S. Department of Commerce, 1973.

   An evaluator's bible! Straightforward, easy to follow. An alphabetical encyclopedia treating every evaluation term you may stumble upon. Sections on evaluation models, functions and targets of evaluation, planning and design, system technologies, variables, measurement approaches and types, technical measurement considerations, reactive concerns, and analysis and interpretation.

3. Anderson, S. B., & Ball, S. The profession and practice of program evaluation. San Francisco, Ca.: Jossey-Bass, Inc., 1978.

   This book answers a broad range of practical questions: What methodological strategies and types of evidence are most useful for various evaluation purposes? How can evaluators ensure greater dissemination and utilization of their findings? How can evaluators deal constructively with their values, preferences and biases? When should evaluators be associated with the program staff and program funding agency and when should they be independent? What competencies do evaluators need and what are the best ways to acquire them?

4. American College Testing Program. Assessment of career development, grades 8-11. Iowa City, Ia: Author, 1973.

   The major objectives of this inventory are to provide information needed in designing effective career education programs tailored to student needs, and to provide a summary of the career development status of student groups for use in assessing the outcomes of career education programs. The test consists of separate subsections on: occupational knowledge; exploratory occupational experiences; preferred job characteristics; career plans; career planning knowledge; career planning activities; and locally developed questions on career development. There are 267 items in varying formats: multiple choice, matching, 3- and 4-point rating scales, checklists, and dichotomous choices.

5. Arterbury, E., et al. The efficacy of career education. Washington, D.C.: National Advisory Council for Career Education, 1976.

   This report presents information about effective programs in career education. The report is organized around four career education thrusts: Career Awareness, Academic Achievement, Career Decision Making, and Other Ways of Assessing Effectiveness. The section on "other ways" lists 14 changes as indicators of success of career education in educational reform (e.g., substantial increase in career education offerings). Each section summarizes findings of other reports and compendiums, bringing together available evidence of effectiveness (evaluation results).

6. Baker, E. L. Formative evaluation in instruction. In W. J. Popham (Ed.), Evaluation in education: Current applications. Berkeley: McCutchan Publishing Corp., 1974, 531-585.

   This excellent chapter contains one of the best available discussions of formative product evaluation. Chapter subheadings include: What is Formative Evaluation For?, Formative Evaluation of Instructional Prototypes, External Data Gathering in Prototype Tryouts, Prototype Data Sources: Summary, Issues in Prototype Testing, Operational Testing, Optional Exercises, and References.

7. Baker, R. L., & Schutz, R. E. (Eds.). Instructional product development. New York: Van Nostrand Reinhold Co., 1971.

   This book is the closest thing to a bible for educational product developers. Chapters cover the whole range of product development, from stating component objectives to operational product field testing. An added benefit is that the book is designed to provide self-tests, and systematic instructional feedback, making it particularly effective for individual study. James Popham's chapter on criterion measures is especially useful.

8. Berdie, D. R., & Anderson, J. F. Questionnaires: Design and use. Metuchen, N.J.: The Scarecrow Press, Inc., 1974.

   The purpose of this book is to help the reader to design and use questionnaires. The authors enumerate and clarify key issues regarding "study-specific" variables. The concepts presented in this book are based on the authors' belief that most criticisms of studies using questionnaires should be directed against poor research design or poorly constructed instruments rather than against the rise of questionnaires per se.

9. Brickell, H. M. Needed: Instruments as good as our eyes. Journal of Career Education, Winter 1976, 2(3), 56-66.

   A case study of the problems encountered in the development, pilot testing, field testing and use of a career development instrument for public school students. This paper provides ample evidence of the problems of matching test items to expected behavior - evaluation instrument to desired outcomes. A humorous, but sobering presentation.

10. Brickell, H. M. The influence of external political factors on the role and methodology of evaluation. Evaluation Comment, 1976, 5(2), 1-6.

    An amusing and perceptive analysis of the problems of evaluators who consistently fail to find evidence of the effectiveness of programs they are hired to evaluate.

11. Bonnet, D. G. Checklist for reporting results of student outcome studies in career education. Crawfordsville, Ind.: New Educational Directions, Inc., 1977.

    The checklist was designed for career education program staff and for evaluators in reviewing draft evaluation reports. It is a guide for ensuring that such reports contain all the information needed by others in interpreting reported results. Nine major items are included: program objectives; program characteristics; target population; measurement tools; test administration; descriptive statistical results; inferential statistical parameters and results; interpretation of results; and final checks.

12. Bonnet, D. G. The design of career education evaluations in 1975-76. Crawfordsville, Ind.: New Educational Directions, Inc., 1977.

    Presents a critique of common evaluation practices as observed in an analysis of the evaluations of projects funded by the Office of Career Education in 1975-76. Also presents unique solutions to common evaluation problems, and ends with a section for project directors on how to get the most from a third part evaluator.

13. Bronson, W. H. California evaluation improvement project. Sacramento, Ca.: California State Department of Education, 1975.

    A model for training practitioners to evaluate their educational programs. Included in the materials are a trainer's manual for use in a workshop setting and participants' materials and exercises. The materials lead participants through evaluation plans and procedures, evaluation design, assessment instruments, data collection, data analysis, evaluation reports, and application of findings.

14. Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago, Ill.: Rand McNally and Co., 1966.

    One of the classics in the field of educational and social program research and evaluation. A must reference resource for anyone attempting to design studies in this field. It provides detailed explanations and illustrations of alternate research and evaluation designs distinguished by their explicit purposes.

15. Capitol Publications. How to evaluate education programs: A monthly guide to methods and ideas that work. Washington, D.C.: Capitol Publications, 1977-78.

A monthly newsletter that treats a different aspect of evaluation in each issue. Topics covered include Choosing a Test; How to Rate and Compare Criterion Referenced Tests; How to Set Evaluation Standards; How to Choose an Evaluation Design; How to Collect Evaluation Information; How to Validate Evaluation Instruments. Practical, easy to follow.

16. Corder, R., & Watkins, R. W. Search for and development of instruments measuring student outcomes of experienced-based career education programs. Princeton, N.J.: Educational Testing Service, 1976.

Reports results of a thorough search for existing instruments that showed promise of being useful in the assessment of anticipated student outcomes. A matrix was developed to show relationships between and among goals and instruments and measuring procedures. Although only the Comprehensive Tests of Basic Skills were accepted by project staff and NIE, the procedures used should be of interest to project personnel concerned with instrument selection and/or development. The analyses of the internal consistency and intercorrelations of various quantitative measures of student outcomes are of particular interest.

17. Datta, L. How to tell a good evaluation plan when you see one and what to do if you don't. Washington, D.C.: National Institute of Education/DHEW, 1978.

As the title indicates, this is a practical discussion which will help practitioners avoid the most common evaluation problems.

18. Ellis, S.; Ellis, N.; & Mason, F. A California study of career centers. Sacramento, Ca.: California State Department of Education, 1975. ERIC/CAPS.

Report of a study of objectives, facilities, equipment, materials, and process and outcome evaluation of a randomly selected sample of career centers in California public schools. This report provides a wealth of information about effective and ineffective practices in planning, establishing, maintaining and evaluating career education facilities.

19. Enderlein, T. A review of career education evaluation studies. Washington, D.C.: U.S. Government Printing Office, 1976.

Summarizes the results of several existing evaluation studies. Reports on findings in four areas: career awareness, career decision making, academic achievement, and other ways of assessing effectiveness. There has been more evidence of effectiveness of programs designed to increase career awareness, than in the other three areas. Identifies programs that have documented evidence of effectiveness.

20. Fagan, B. M., & Horst, D. P. Selecting a norm-referenced test (Technical Paper No. 4). Mountain View, Ca.: RMC Research Corporation, 1976.

Written simply and concisely, this booklet outlines four guidelines for selecting norm-referenced tests. These suggestions are intended for practitioners who face the difficult task of identifying appropriate instruments for assessing student achievement.

21. Flanagan, J. C. Use of comprehensive rationales. Educational and Psychological Measurement, 1951, 11, 151-155.

A brief methodological note primarily of historical interest. Its basic thesis: make sure you know how the behavior you are actually measuring relates to the trait or phenomenon you think you are measuring.

22. Flanagan, J. C. Measuring human performance. Palo Alto, Ca.: American Institutes for Research, 1974.

This text is especially valuable if you are interested in attitude or value measurement. It suggests that too much dependence in this area is placed on self-report ratings and scales and not enough on the observation of behavior - a valid indicator of attitude or value.

23. Fullan, M., & Pomfret, A. Research on curriculum and instruction implementation. Review of Educational Research, 1977, 47(1), 335-397.

A detailed article that really gets into the whole topic of the implementation of educational innovations. The authors urge evaluators to investigate more thoroughly the actual implementation process including (1) the extent to which the innovation was operationalized the way it was planned, and (2) the nature and degree of adaptations made during that implementation process.

24. Richard A. Gibboney Associates, Inc. The Career Intern Program: Final report. Washington, D.C.: National Institute of Education, 1977.

In two volumes, this comprehensive report provides a model for program evaluation as evaluation relates to the various parts of the development and implementation processes. Although the documents are heavy with reports of collected data, what the authors have done with the data is worthy of note. An excellent example of a comprehensive project report, with good display of evaluation design, processes, and data collection and analysis.

25. Gilchrist, R. S., & Roberts, B. Curriculum development: A humanized systems approach. Belmont, Ca.: Fearon Publishing Co., 1974.

Chapter 6, "Evaluation as Communication," contains an excellent discussion of the decision making and communication needs involved in educational evaluation. It deals with process and product evaluation, and, to a lesser degree, with data gathering and measurement techniques.

26. Gill, W. B., & Luke, A. W., Facilities handbook for career education. Washington, D.C.: National Institute of Education, 1976.

Although this document focuses on planning and implementation of facilities, together with descriptions of specific facilities for career education, the practitioner who wishes to evaluate a planned or existing facility will find many points worth pursuing in an evaluation. An extensive bibliography is presented.

27. Gilmartin, K. J., Rossi, R. J., & Russ-Eft, D. Program impact evaluation paradigm for VISTA. Palo Alto, Ca.: American Institutes for Research, 1977.

This is one of the volumes created in a project to provide evaluation models for the full-time domestic volunteer programs sponsored by the ACTION agency. A set of concepts is presented, defined, and used to provide a "common yardstick" for all of the various services and activities that VISTA sponsors or supports. The index of program impact selected was the constituency strength of a VISTA community, that is, the extent to which the communities affected by VISTA programs engage in self-directing behaviors. Investment behavior, behavior involving the expending of personal resources for expected positive return, and opportunity, the perceived costs of investments relative to expected returns, were key components of this index. A comprehensive list of indicators for the relevant variables is supplied.

28. Gronlund, N. E. Preparing criterion-referenced tests for classroom instruction. New York: The Macmillan Co., 1973.

This book is intended as a practical guide for the preparation and use of criterion-referenced tests in classroom instruction. Topics include the following: the nature of criterion-referenced testing; the role of criterion-referenced testing at two levels of learning, mastery and development; procedures for planning the test; writing test items; using and appraising the test; a checklist for evaluating criterion-referenced tests; and a list of references.

29. Gushee, T.,& Organt, G. J. Evaluation of student progress: Program of prevocational education, Prince George's County, Maryland Public Schools. Upper Marlboro, Maryland: Prince George's County Public Schools, Career Education Department, 1976.

This report, which documents the effectiveness of the prevocational education component of Prince George's County's career education program, provides some good illustrations of how to display quantitative data in project reports.

30. Hamilton, J., & Mitchell, A. Identification of Evaluated Exemplary Career Education Activities: Final Report. Palo Alto, Ca.: American Institutes for Research, 1978.

This report describes criteria for analyzing evaluation reports of career education activities, and identifies nominated projects as well as those that met the project's criteria and were submitted to the Joint Dissemination Review Panel for review and approval. Criteria provide guidelines for evaluators; study of these guidelines may help evaluators avoid some common problems and design evaluations that will provide evidence of the effectiveness of their programs.

31. Harsh, J. R. The forest, trees, branches, leaves, revisited - Norm, domain, objective and criterion-referenced assessments for educational assessment and evaluation. Association for Measurement and Evaluation in Guidance, Monograph No. 1. Fullerton, Ca.: California Personnel and Guidance Association, 1974.

An excellent discussion about the various types of measurement instruments, and the appropriate use of each. A must for anyone who is not clear about the distinctions between and among norm, domain, objective, and criterion-referenced assessment.

32. High, S. C., Jr. Career education: What proof do we have that it works? Paper presented at the Commissioner's National Conference on Career Education, Houston, November 1976.

Presents examples of the kind of evidence that can be arrayed regarding the ability of career education programs to attain six of the nine learner goals proposed by the Office of Career Education.

33. Horst, D. P., & Fagan, B. M. Common evaluation hazards (Technical Paper No. 14). Mountain View, Ca.: RMC Research Corporation, 1976.

A practical discussion of potential problems in sampling, data collection, selection of evaluation design, instrument selection/development. Helps the evaluator plan a problem-free evaluation.

34. Hoyt, K. B. Perspectives on the problem of evaluation in career education. Washington, D.C.: U.S. Office of Education, 1974.

This report is a synthesis of discussions by experts and practitioners at two mini-conferences. It treats problems of instrumentation and evaluation and gives examples of approaches to evaluation in career education. It discusses use of the Office of Career Education's nine learner outcomes in evaluation efforts.

35. Hoyt, K. B. A primer for career education. Washington, D.C.: U.S. Government Printing Office, 1976.

Probably the most basic and important of Hoyt's publications, the Primer presents straightforward, uncompromising definitions and roles for a variety of "actors" in career education. Although the section on evaluating career education is relatively brief, the entire document gives focus to the purposes of career education, and thereby defines parameters for evaluation.

36. Instructional Objectives Exchange. Illustrative criterion-referenced test specifications: Career education (K-12). Los Angeles, Ca.: Author, 1976.

This excellent catalog outlines a strategy for developing criterion-referenced tests in career education. It also provides examples of test items keyed to specific career education objectives.

37. Jones, G. B., Wolff, J. M., Quirk, T. J., Unruh, W. R., & Jung, S. M. Evaluation: A training model for school personnel utilization. Palo Alto, Ca.: American Institutes for Research, 1974.

Five competency-based modules, each addressing one of the following:
1. Formulating a General Evaluation Approach
2. Specifying Product and Process Objectives
3. Measuring Outcomes
4. Analyzing and Synthesizing Data
5. Using and Reporting Results.

Each is a detailed, easy to follow guide with examples.

38. Jung, S. M. Evaluating the Work-Education Consortium Project: An overview of issues and options. Washington, D.C.: National Manpower Institute, 1977.

Presents a general evaluation framework and discusses the functions of evaluation and the evaluator's role. Particularly helpful are the forms and checklists provided.

39. Kester, R. J. Using systematic observation techniques in evaluating career education. Columbus, Ohio: The National Center for Research in Vocational Education, The Ohio State University, 1978.

This handbook establishes a basic rationale for the use of systematic observation as a legitimate technique for use in evaluating career education. The three techniques of observation, interview, and unobtrusive data collection are discussed. Additional sections are provided on things to consider when using systematic observation, how to organize, analyze, and justify the use of systematic observation data and enhancing your observation skills.

40. Kiefert, J., & Maser, A. Program/project evaluation guide. Seattle, Wa.: University of Washington Press, 1977.

This workbook presents a model for evaluating educational programs, with simple, easy to follow guidelines for simplifying those components of evaluation that many practitioners assume are difficult. Particularly helpful sections are those on evaluation data requirements, evaluation design, data analysis and processing, and program monitoring.

41. Klein, S. B. Toward consensus on minimal criteria for educational products. Paper presented at AERA Annual Meeting, San Francisco, California, 1976.

Discusses criteria, decision makers, and types of educational products. Decisions concerning product development, product screening and product selection are treated. Presents an interesting discussion on the difficulty of achieving consensus on criteria for assessment of educational products. Since selection/development of educational products must be treated in evaluating programs using such products, this paper is suggested as an appropriate source of insight.

42. Malak, S. Assessing experiential learning in career education. Columbus, Ohio: The National Center for Research in Vocational Education, The Ohio State University, 1978.

This handbook presents the experiences of a number of persons actively involved in experiential programs and their assessment at the secondary and postsecondary level. Additionally, it provides examples of alternative evaluation techniques and guidelines to follow in assessing experiential learning in career education.

43. McCaslin, N. L., Gross, C. J., & Walker, J. P. Career education measures: A compendium of evaluation instruments. Columbus, Ohio: Ohio State University, Center for Vocational Education, 1977.

Identification and selection of evaluation measures and materials for career education programs continue to be major problems for project staffs. This compendium of career education evaluation measures presents a good discussion of the potential questions to be asked when selecting career education measures, as well as a section on abstracts of educational measures. Each abstract presents title, author, availability, instrument description, example item, administration information, and availability of technical data. This is the only compendium of career education measures that is reasonably complete and current.

44. McCaslin, N. L., & Adams, K. A. Managing and evaluating career education. Columbus, Ohio: The National Center for Research in Vocational Education, The Ohio State University, 1977.

This document presents evaluation techniques for each of the following: planning, organizing, directing, staffing, and controlling. Managers are given some very practical techniques for participating in each of these functions. The specific tasks for each are presented, with background, definitions, examples and resources. A helpful document for managers.

45. McCaslin, N. L., & Walker, J. P. A guide for improving locally developed career education measures. Columbus, Ohio: Ohio State University, Center for Vocational Education, 1978.

Discusses some of the problems encountered when attempting to improve the planning and designing of career education instruments, and suggests ways for avoiding these problems. Discusses test reliability and test validity, and offers some useful guidelines for gathering reliability and validity data on locally-constructed tests.

46. McLaughlin, D. H. Career education in the public schools: 1974-75: A national survey (HEW Publication No. (OE) 76-00507). Washington, D.C.: U.S. Government Printing Office, 1976.

Presents results of: (1) local school district career education activities; (2) state level activities; (3) teacher college activities; and (4) materials for career education. Estimates frequency of different activities, compares programs in different types of districts and regions, and identifies the involvement of various types of educators. Reports that increased community involvement emerged as an essential component of successful career education programs.

47. McLaughlin, D. H., Gilmartin, K. J., & Rossi, R. J. Controversies in the evaluation of compensatory education. Palo Alto, Ca.: American Institutes for Research, 1977.

Discusses the role of evaluation in educational program development and identifies issues of evaluation design, sampling, measurement, and analysis. Discusses methodological issues, their importance for interpreting evaluation results, and alternative resolutions. Makes 27 recommendations for improving evaluation methodology.

48. National Assessment of Educational Progress. Career and occupational development objectives: Second assessment. Denver, Colorado. National Assessment of Educational Progress, 1977.

Describes the development of the objectives to be used in the second national assessment of career and occupational development. Lists objectives in two categories: knowledge, abilities, and attitudes relevant to career decisions, and knowledge, abilities, and attitudes necessary for success in a career or occupation. A good source of objectives that may coincide with the objectives of local programs. Contains 51 test items for 60 objectives related to 7 of the 9 OCE outcomes. Test items are available from NAEP.

49. New Educational Directions, Inc. What does career education do for kids?: A synthesis of 1975-76 evaluation results. Crawfordsville, Ind.: Author, 1977.

A final report of a nationwide project conducted for the United States Office of Education (USOE). This study investigated the evaluation designs, procedures, results, and conclusions of evaluations conducted on a set of USOE-supported career education projects. This report summarizes some of the commonalities discovered across those project evaluations.

50. New Educational Directions, Inc. Checklist for reporting results of student outcome studies in career education. Crawfordsville, Ind.: Author, 1977.

In an attempt to help evaluators improve their reports of student outcome studies evaluating career education activities, this booklet outlines a list of guidelines. These suggestions help report writers determine what should be done, when, and how.

51. Owens, T. R., & Evans, W. D. Program evaluation skills for busy administrators. Portland, Ore.: Northwest Regional Educational Laboratory, 1977.

Discusses purposes of evaluation and presents guidelines for describing a program or activity, establishing evaluation guidelines, preparing an evaluation plan, implementing the evaluation, and reporting evaluation findings. A practical, how-to guide.

52. Peat, Marwick, Mitchell, & Co. A systematic approach to evaluating career education materials at the local level. Washington, D.C.: U.S. Office of Education, 1974.

This publication presents a systematic approach to evaluating career education materials. The authors developed an assessment instrument that is presented in the publication, along with suggestions for its use. Materials are assessed according to quality, conditions for use, content bias, and research data provided.

53. Plakos, M.; Plakos, J.; & Babcock, R. Program evaluation workbook, California Evaluation Improvement Project. Sacramento, Ca.: California State Department of Education, 1975.

This workbook helps the evaluator define the purposes of evaluation; review needs assessment information; review program goals and objectives; identify evaluation data requirements; determine the evaluation design; identify assessment instruments; develop data collection procedures; select data analysis techniques; plan for the processing of data; develop program monitoring procedures; develop a reporting plan; develop an evaluation timeline; identify required resources; and chart a plan. This is one part of a total evaluation program that constitutes the California State Model.

54. Platt, J. R. Strong inference. Science, 1964, 146, 347-352.

A pioneering and oft-cited article on the scientific method. Platt convincingly argues that most hypotheses in science are not proved; rather, plausible alternative hypotheses are disproved, strengthening the support for the original hypothesis. Platt's admonition is worth remembering: what experiment would disprove your hypothesis?

55. Research and Evaluation Unit. 1976-77 final report of career education (Project EPIC). Report #14. Louisville, Ky.: Jefferson County Public Schools, 1977.

The final report of an evaluation conducted in Project EPIC during 1976-77. This evaluation, performed by the Research and Evaluation Unit of the Jefferson County (Kentucky) Public Schools, documented student growth promoted by this career education program.

56. San Diego County Department of Education. Annotated directory of career tests and resources. San Diego, Ca.: San Diego County Department of Education, 1975.

Developed during establishment of a model state career education resource center, this directory presents tests and resources in several categories, describes them, and suggests appropriate uses. Only commercially available tests and materials are listed.

57. Schutz, R. E. Programmatic instructional development. Paper delivered at the annual meeting of the American Educational Research Association, Minneapolis, Minn., March 1970.

A forthright statement of the trials, tribulations, and prospects for systematic, research-based development of educational products, by one of the pioneers in the field.

58. Shaycoft, M. F. Guide to the development, evaluation, and use of criterion-referenced tests. Cambridge, Mass.: Garland Publishing Inc., forthcoming.

The title indicates this publication's main focus, criterion-referenced tests. The viewpoint expressed, however, is that criterion-referenced tests and norm-referenced tests are more alike than is usually supposed. In fact they may even be the same instrument.

59. Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

This book presents the nonparametric techniques of hypothesis testing in a usable form. For each example of the different kinds of statistical tests discussed, the author describes the rationale underlying the test, explains its computation, gives examples of its application, and compares the test to its parametric equivalent. Many statistical tables are included as an appendix.

60. Sorenson, G. Evaluation for the improvement of instructional programs: Some practical steps. Evaluation Comment, 1971, 2(4), 13-17.

The Evaluation Comment newsletter is published by UCLA's Center for the Study of Evaluation. This article from the January 1971 issue is excellent because of the breadth of Sorenson's coverage on the topic of evaluation in product development. It includes a "Formative Evaluation Checklist."

61. Tallmadge, G. K. The Joint Dissemination Review Panel: Ideabook. Mountain View, Ca.: RMC Research Corporation, 1977.

Discusses the following hazards in evaluation:

1. Claiming much, providing evidence of little
2. Selecting measures not logically related to the intervention
3. Use of grade-equivalent scores
4. Use of a single set of test scores for both selecting and pretesting participants
5. Use of comparisons with inappropriate test dates for obtaining information
6. Use of inappropriate levels of tests
7. Missing data
8. Use of noncomparable treatment and control groups
9. Use of inappropriate statistical adjustments with nonequivalent control groups
10. Constructing a matched control group after the treatment group has been selected
11. Careless collection of data
12. Use of different instruments for pretesting and posttesting
13. Use of inappropriate formulas to generate no-treatment expectations
14. Mistaken attribution of causality

62. Tallmadge, G. K. Criterion-referenced tests (Technical Paper No. 11). Mountain View, Ca.: RMC Research Corporation, 1976.

A brief booklet that delves into key issues related to the selection of criterion-referenced and norm-referenced instruments for assessing student achievement. It especially addresses the conditions under which criterion-referenced tests should and should not be used.

63. Tallmadge, G. K., & Wood, C. T. Characteristics of eight commonly used, nationally normed tests (Technical Paper No. 5). Mountain View, Ca.: RMC Research Corporation, 1976.

A booklet that provides concise annotations on eight norm-referenced instruments for assessing student achievement in the cognitive area. These tests are well known and are supported by national norms.

64. Tallmadge, G. K., & Wood, C. T. User's guide: ESEA Title I evaluation and reporting system. Mountain View, Ca.: RMC Research Corporation, 1976.

Presents several designs for evaluating Title I programs. Each design suggests use of a specific type of comparison standard for measuring student gains in basic skills.

65. Tallmadge, G. K., & Wood, C. T. Proposed Title I evaluation forms and instructions. Mountain View, Ca.: RMC Research Corporation, 1976.

Forms and instructions for several designs for evaluating Title I programs.

66. Texas Partners in Career Education. Career education learning system. Arlington, Texas: Texas Education Agency, 1976.

The project developed 17 tests, including 273 items, for 63 basic learner outcomes covering 4 of the 9 OCE learner outcomes.

67. Tuckman, B., & Carducci, J. Evaluating career education: A review and model. Washington, D.C.: National Institute of Education, 1974.

This paper examines and reviews evaluations of career education programs in the literature current at that time, and suggests a model or paradigm for conducting such evaluations. The paper reviews experimental and quasi-experimental evaluative studies; descriptive evaluations; and case studies. The authors offer a useful framework for assessing career education programs, and present a good discussion of dependent and independent variables, as well as control variables as these affect evaluation. Both descriptive and inferential data analyses are treated.

68. Upton, A., Lowrey, B., Mitchell, A., & Varenhorst, B. A planning model for developing a career guidance curriculum. Fullerton, Ca.: California Personnel and Guidance Association, Monograph No. 12, 1978.

This planning model, developed by the California State Department of Education, Office of Pupil Personnel Services, has a chapter on evaluation that is so concise as to suggest an outline, yet with enough definition to help the evaluator know whether to seek additional information elsewhere.

69. Van Dalen, D. B., & Meyer, W. J. Understanding educational research (2nd. ed.). New York: McGraw-Hill, 1966, pp. 301-325.

This section presents simple straightforward instruction on the construction of paper and pencil instruments including questionnaires, tests, inventories, rating scales, checklists, sociometrics, etc.; performance tests, behavioral observations, including checklists and schedules, time sampling, etc.; and interviews, including individual and group, structured and unstructured. Only 25 pages, but full of helpful information.

70. Webb, F., Campbell, D., Schwartz, R., & Sechrest, L. Unobtrusive measures: Nonreactive research in the social sciences. Chicago: Rand McNally, 1966.

In an attempt to broaden the perspectives of social science researchers, this book encourages the use of innovative research methods, but purposely excludes the most common methods of interviewing and administering questionnaires. Physical traces, archives, and observations are the three major categories of research methods which are explored. The book contains an extensive bibliography (28 pp.).

71. Wellman, F. E., & Moore, E. J. Pupil personnel services: A handbook for program development and evaluation. Washington, D.C.: U.S. Office of Education, 1975.

This handbook was developed as a guide for program development and evaluation of pupil personnel services. It discusses evaluation-based program development; evaluation design (formulation of hypotheses, specification of a comparison base, identification of participants, determination of analysis procedures); evaluation structure (instrumentation, scheduling data collection, staffing for data collection and processing, budgeting for evaluation), questions to be asked, evaluation implementation (collection of outcome and process data, analysis of evaluation data), and reporting evaluation results. The appendices present examples of needs assessment forms and procedures, outcome instruments and procedures, and process monitoring and reporting procedures.

72. Wise, L. L. The fight against attrition in longitudinal research. Palo Alto, Ca.: American Institutes for Research, 1977.

A detailed analysis of the causes of participant attrition in longitudinal studies and ways to help reduce it and deal with it statistically.

73. Young, M. B., & Schuh, R. G. Evaluation and educational decision making: A functional guide to evaluating career education. Washington, D.C.: Development Associates, 1975.

A comprehensive model for evaluation of career education activities. In 12 carefully planned and well explained steps, the guide carries the evaluator through the total evaluation process. Interrelationship of steps is treated, and possible problems, solutions, and rationales are presented. One of the best and most complete documents available to the practitioner. Casual readers may find the terminology confusing, so this guide is best used as a reference work for evaluation topics about which there is a particular question.

APPENDIX A

Ten Evaluated Career Education Activities

The rationale chapter of this document states that the development of this Handbook was pursuant to a contract with the Office of Planning, Budgeting and Evaluation, U.S. Office of Education. The purpose of the contract was to identify evaluated, exemplary career education activities, K through 12. In all, 394 activities were nominated by career education coordinators, evaluators, educators and practitioners from the 50 states. Of those nominated, 257 felt that they met the basic criteria for inclusion in the search (evidence of effectiveness, evaluated since 1974, and a written evaluation report available) and submitted their evaluation reports for analysis.

Criteria established by the project staff for use in the analysis included evidence of adequate rigor in performance of each of the steps included in the appropriate evaluation model in this Handbook. Criteria also reflected those employed by the Joint Dissemination Review Panel (JDRP).[1] Reports of the evaluation of federally supported educational products and practices may be submitted to the JDRP for review. Those presenting evidence of effectiveness are approved as exemplary, and are eligible for national dissemination.

Of the 257 activities analyzed for this project, few presented strong evidence of having met all criteria established by the project staff. Ten were selected for submission to the JDRP, and of these seven passed the review. All of these projects were evaluated between July, 1974, and August, 1977 when this project was initiated. The fact that seven were approved by the JDRP is evidence that even before there were many models and guides available, local education agencies were following sound practices in evaluating their career education activities. To convince you, the practitioner, that real live people have conducted creditable evaluations, we are listing the 10 submitted to the JDRP, and presenting a short abstract of each. Those marked with an asterisk were approved by JDRP. The others may be resubmitted after additional documentation is available in areas questioned by the Panel.

---

[1]The Joint Dissemination Review Panel (JDRP) was established by the Education Division of the Department of Health, Education, and Welfare in 1972. The Joint U.S. Office of Education (OE) - National Institute of Education (NIE) Panel currently has 22 members, eleven from each agency, appointed respectively by the Commissioner of Education and the Director of NIE.

PROJECT OVERVIEWS

* Akron Career Development Program
Office of Career Education Programs
Akron, Ohio

The goals of Akron's Career Development Program are to increase students' self-awareness and self-esteem, knowledge of the world of work, and skills in decision-making. Seven areas of objectives have been developed based on these goals: the self; the individual and the environment; education and training; the world of work; economics; employability and work adjustment; and decision-making.

Equal numbers of randomly selected program and non-program students were selected in the spring of 1976 for testing on a posttest-only basis. Program students' performance was compared with that of non-program students.

* Project CAP (Career Awareness Program)
Boston Mountains Educational Cooperative for Federal Programs
Greenland, Arkansas

The goals of Project CAP are to provide awareness of adult occupations among K-8 level students, along with regular instruction in basic academic areas. Special emphasis is placed on showing students that the studying they are doing in school is important and related to work they will encounter outside of school.

The impact upon students of Project CAP activities implemented in 1976-77 was assessed by testing comparable groups of treatment and control students in February and March 1977.

* Project CERES (Career Education Responsive to Every Student)
Ceres Unified School District
Ceres, California

The project has the dual goals of infusing career education concepts into the elementary and secondary curriculum and of developing an articulated curriculum that provides for student attainment of career education goals.

The impact of Project CERES activities on students was evaluated in 1973-74 by utilizing a pre-post treatment and control group design.

* Project CDCC (Career Development Centered Curriculum)
Coloma Community School District
Coloma, Michigan

The goal of Project CDCC is to provide career development instruction to rural students at the elementary level (K-6). This instruction is designed especially to increase students' awareness of life roles, to help them formulate realistic self concepts, and to foster the development and use of decision-making skills.

The impact of Project CDCC activities upon students was assessed in 1975-76 by testing comparable groups of treatment and control students at the beginning and again at the end of each instructional unit.

Project EPIC (Educational Progress In Careers)
EPIC, Inc.
Fort Lauderdale, Florida

The three goals were to: (1) incorporate employers' and other business leaders' concerns into the concepts of career education; (2) collaboratively develop an instrument to assess whether career education has been implemented according to those concepts; and (3) develop valid and reliable instruments to assess student outcomes for those students who have been involved in career education as defined by the concepts developed in the first goal.

The instruments that were developed under goals two and three above were required to pass tests of content validity, usability, construct validity, and reliability. Content validity was cross checked by both state and local educational and occupational experts working in task force groups. Usability was established by teacher-review and by pilot testing the instruments. Construct validity and reliability were assessed in a formal field test conducted in career education and control classes in three Florida counties.

* Project Equality
Highline Public Schools
Seattle, Washington

The goal of Project Equality is to expand students' perceptions of occupations open to females and those open to males. The specific objectives to attain that goal have dealt primarily with the acquisition and development of media and materials that enhance the availability of choice for all students in the elementary classroom and reflect the changing definition of what "women's work" is.

The impact of Project Equality activities upon students was assessed in 1977-78 by testing comparable groups of treatment and control students.

Project EPIC (Educational Preparation for Involvement in Careers)
Jefferson County Public Schools
Louisville, Kentucky

The goal of Project EPIC is to provide cumulative experiences across grades K-12 that are designed especially to increase low-income students' self-awareness and basic academic skills. Special emphasis is placed on aiding students to become aware of a variety of careers and on teaching decision-making skills.

The impact of Project EPIC activities upon students was assessed in 1976-77 by testing treatment students at the beginning (pretest) and again at the end of the school year (posttest). Student scores were compared with a national average or with predicted posttest scores.

* Project MATCH (Matching Attitudes and Talents to Career Horizons)
Ontario-Montclair School District
Ontario, California

Project MATCH is designed to infuse a career education component into the regular curriculum of a K-8 school district program. Goals have been formulated in ten areas of career education, including career awareness, self-awareness, consumer competencies, and career planning and decision-making.

The evaluation design utilized pre- and posttests on intact treatment and control classes. These classes were randomly selected, one class per grade, from schools assigned to the treatment and control conditions. Student attainment of project objectives was assessed by four project-developed career education tests with established reliability and validity.

* Pima County Developmental Career Guidance Project
Tucson, Arizona

The ultimate goal of this project is to help all K-12 students in the county grow in the areas of self-awareness, educational awareness, career awareness, economic awareness, decision-making, beginning competency, employability skills, and appreciations and attitudes.

The impact of this project on student growth was assessed in 1974-75 by testing comparable groups of treatment and control students on a posttest only basis. The grades tested were 4-12.

Fostering a Team Approach to Career Education
Prince George's County Public Schools
Upper Marlboro, Maryland

The goal of this project was to implement in each junior high school a successful team approach to career education. Successful implementation of the team approach was determined by assessing its impact upon students.

The impact of the team approach upon students was assessed in 1975-76 by testing treatment and control students at the beginning (pretest) and again at the end of the school year (posttest).

APPENDIX B

Examples of Promising Activity Evaluations

As the project staff analyzed the 257 evaluation reports that were submitted for review, they found many that were not yet ready for JDRP review and possible dissemination, but that showed promise of providing evidence of their effectiveness. A few of these are listed here to give you an idea of the kinds of career education activities that are being implemented and evaluated.

- Region XII Career Education Program, Kentucky Valley Educational Cooperative (Hazard, Kentucky). This is a project to communicate successful career education methods and procedures in rural Appalachia K-12 settings. They have developed and disseminated 3000 copies of Career Education: The Region XII Approach, 17 "how to" documents, and Career Education Concepts and Skills: A Guide for Teachers in Rural Areas. The project has a strong staff development component. Student gains are shown in the areas of academic skills, career awareness, and attitudes toward school and work.

- Huntington Beach Union High School District (California). Operational Model for Career Development and Vocational Preparation. This project identified specific competencies students needed in individual career development, developed a sequential program to deliver the competencies in an articulated K-12 model, and incorporated the model into existing resources and strategies. This competency-based model encompasses three functional levels: awareness, internalization, and action, and is keyed to the California State model for Career Guidance Curriculum.

- Career Guidance Institutes (Maryland State Department of Education). This is a comprehensive staff development project, focusing on counselor development of guidance competencies. A one-group design with repeated measures was used.

- Career Education Measurement Series (Texas Education Agency). A measurement and diagnostic system designed to support instruction in career education. There is a battery of 17 tests for students in grades 7-12 consisting of one survey test and 16 category tests.

- The Thaler System (Gilman, Vermont). The name derives from the Greek work thalos, which means dollar, or unit of money. The students have developed an economic system. There are impressive trend data in that math performance on standardized achievement tests is up 20%.

- Occupational Counseling for Physically Handicapped Students (New York). This is a tutorial system using graduate student interns. The evaluation design utilizes treatment and comparison groups. Significant differences were found when treatment students were administered one-half of the CMI.

- St. Tamany Parish Career Education Program (Louisiana). This is a K-12 infusion program. Student performance data were collected at the end of the 1977-78 school year; preliminary analyses look promising.

- Me in the Future (Boulder, Colorado). This is a project to produce and evaluate model science materials for educable mentally handicapped secondary students. The materials cover six occupational clusters (e.g., agri-business, construction, personal services); they were field tested in 12 sites nationwide including inner-city, urban, suburban, and rural settings. The project is in the formative evaluation stage.

- Career Intern Program (Livonia, Michigan). This project is for students in grades 11 and 12. There were 11 participants in 1976-77 and 60 participants in 1977-78. With the limited sample from the first year, significant differences were obtained on the CMI subtest Choosing a Job.

- From Dreams to Reality: A Career Awareness Exploration Project (New York). This project was developed by the Girl Scouts and supported by OCE. It operated with many Girl Scout troops throughout the country. There is a good evaluation design.

- Total Opportunities Through Action Learning (Hillsboro, Oregon). This is an Experience Based Career Education (EBCE) project based on the Northwest Regional Educational Laboratory (NWREL) model. Significant achievement results were obtained on the following: CTBS Reading Test; Occupation Selections scale of the Self-Directed Search; and the Growth in Self-Reliance area of the Student Opinion Scale.

- Willoughby-Eastlake (Ohio). This is an infusion project comparable to the Ohio State Career Education model and to the Akron project. Staff have developed a Career Orientation Curriculum Guide. A fire in September 1976 destroyed some evaluation material so new evidence of effectiveness is being generated.

- Jobs: Seeking, Finding, Getting (Maryland State Department of Education). This is a project to develop an instructional TV series. Sixteen 20-minute telelessons are being developed along with a teacher's manual and student handbook. The series is scheduled to be aired in October 1978.